

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 20, 2005, 00:41:37, Search time 10034.3 Seconds  
(Without alignments)  
11541.209 Million cell updates/sec

Title: US-10-791-017A-1

Perfect score: 2390  
Sequence: 1 agagggagcgcgcgcgttgag.....taaaaaaaaaaaaaa 2390

Scoring table: IDENTITY\_NUC  
Gapop 10.0, Gapext 1.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database:

GenBank: 1: gb\_ba: 2: gb\_hg: 3: gb\_in: 4: gb\_cm: 5: gb\_ov: 6: gb\_pat: 7: gb\_ph: 8: gb\_pl: 9: gb\_pr: 10: gb\_ro: 11: gb\_stb: 12: gb\_sy: 13: gb\_un: 14: gb\_vl:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2390	100.0	2390	6	C0867364 Sequence
2	2388.4	99.9	2390	6	AX411125 Sequence
3	2388.4	99.9	2390	6	HSEWS
4	2369.4	99.1	2371	6	A36460 Sequence 1
5	2369.4	99.1	2371	6	AR080100 Sequence
6	2330	97.1	2364	6	BC011048 Homo sapi
7	2279.2	95.4	2336	6	C0730352 Sequence
8	2176	91.0	2189	9	CR456490 Homo sapi
9	2165	90.6	2182	9	BC004817 Homo sapi
10	2165	89.4	2164	9	BC072442 Homo sapi
11	2123	88.8	2189	9	AK056309 Homo sapi
12	2038.4	85.3	193559	2	AC137500 Homo sapi
13	2038.4	85.1	155815	2	AC018774 Homo sapi
14	2033.6	85.1	180718	2	AL5596087 Human DNA
15	2033.6	85.1	182501	2	AC011221 Human sapi
16	2015	84.3	2396	5	BC083960 Xenopus 1
17	2000	83.7	132906	5	AC109802 Canis fam
18	2000	83.7	163104	2	AC110672 Canis fam
19	1969.4	82.4	1971	12	BT007796 Synthetic

20	1927.4	80.6	2188	6	AX105537
21	1927.4	80.6	2188	10	MHEWS
22	1926.4	80.6	2189	10	BC068226
23	1808.4	75.7	1988	6	AX714390
24	1808.4	75.7	1988	9	AK056681 Homo sapi
25	1782.8	74.6	2026	6	C0850483
26	1782.8	74.6	2026	6	AK127624 Homo sapi
27	1710.8	71.6	226400	2	AC128482
28	1710.8	71.6	235241	2	AC098231
29	1710.8	71.6	263925	2	AC106522
30	1589.4	66.5	221444	2	AC113313
31	1579.8	66.1	247757	2	AC121282
32	1428	59.7	1783	6	C0721057
33	1423.8	59.6	2440	5	AJ719366
34	1353.2	56.6	7293	9	HSN08920
35	1306.8	54.7	221924	2	AC105889
36	1302.4	54.5	275105	2	AC118907
37	1273.8	53.3	270337	2	AC099212
38	1133.4	47.4	2473	9	AK026270
39	1014.8	42.5	1303	9	BC000527
40	979	41.0	1864	9	AF254086
41	979	41.0	2219	9	AF254087
42	979	41.0	2252	9	AF254088
43	848	35.5	2665	5	BC063928
44	836.4	35.0	2107	9	HSU35622
45	834	34.9	2683	5	BC075120

## ALIGNMENTS

RESULT 1  
LOCUS C0867364 2390 bp DNA linear PAT 13-SEP-2004  
DEFINITION Sequence 1 from Patent EP1455190.  
ACCESSION C0867364  
VERSION C0867364.1 GI:51997589  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
REFERENCE 1 Obendorf, M. and Wolf, S.  
AUTHORS Modulation of the activity of nuclear receptors via EMS  
TITLE Patent: EP 1455190-A 1 08-SEP-2004;  
JOURNAL Schering Aktiengesellschaft (DE)  
FEATURES  
source location/Qualifiers  
1. 2390  
/organism="Homo sapiens"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:9606"  
44..2014  
/note="unassigned protein product"  
/codon\_start=1  
/protein\_id="CAH33891.1"  
/db\_xref="GI:51997590"  
/translation="MASTVSTYSQAAGQYSAYTAOPTGCAVATTQAAVGGQSYGY  
GOPTVSYQAOTTAATVGGTAVATSGOPTATVTPAQAASOPVGGTAVDTT  
ATVTTQASVTAQSAVAGTGPADVSDNSALVYGLNDSVTLADLAFKCGVAKKRP  
PSLGGQSTVSPVQPGSTFPMQPVAPAPSPPTTSSTPTSTYDQSSYRQHPGSP  
STGQSSYTGQSSYVQGPSTYPTGTSYSAVSOYSSQSSSYGQSSYRQHPGSP  
VYGGSGSGSPGPNRSGPDRGSGPDRGSGPDRGSGPDRGSGPDRGSGPDRGSGP  
KCGPDSDEGPDLDLGPVDPDSDNSALVYGLNDSVTLADLAFKCGVAKKRP  
TGQPMHLYLDKETGPKGDATVSYEDPTAKAAVEWFGDFQSGKLVSLARKKP  
NMSMRGALPBRGSGPDRGSGPDRGSGPDRGSGPDRGSGPDRGSGPDRGSGP  
PBGCGNVQHRADNOCNPGNCGNCPARTENCKAPVDBFLPPFPFPBGDRRG  
GPGKRGKGLNDRGSGPDRGSGPDRGSGPDRGSGPDRGSGPDRGSGPDRGSGP  
LMEQWGRGRGSGPDRGSGPDRGSGPDRGSGPDRGSGPDRGSGPDRGSGP

ORIGIN  
Query Match 100.0%; Score 2390; DB 6; Length 2390;  
Best Local Similarity 100.0%; Pred. No. 0;









QY 181 GCCCATGATGTGAGCTATACCCAGGCTCAGACCACTGCAACCTATGGGCAAGCCGCTTA 240  
DB 181 GCCCATGATGTGAGCTATACCCAGGCTCAGACCACTGCAACCTATGGGCAAGCCGCTTA 240  
QY 241 TGCATCTTCTTATGGAAGGCTCCCTGCTTACTACTCAACTGCTCCGCCAGGATTA 300  
DB 241 TGCATCTTCTTATGGAAGGCTCCCTGCTTACTACTCAACTGCTCCGCCAGGATTA 300  
QY 301 CAGCCAGCTGTCCAGGGGTTATGGCACTGTGCTTATGATACCACTGCTCAAGTCAAC 360  
DB 301 CAGCCAGCTGTCCAGGGGTTATGGCACTGTGCTTATGATACCACTGCTCAAGTCAAC 360  
QY 361 CAGCCAGCTGTCCAGGGGTTATGGCACTGTGCTTATGATACCACTGCTCAAGTCAAC 420  
DB 361 CAGCCAGCTGTCCAGGGGTTATGGCACTGTGCTTATGATACCACTGCTCAAGTCAAC 420  
QY 421 CTATGGGCAAGGCAAGCACTGCACTGCACTCAAGACCGGAGGATGAAACAAAGCCAC 480  
DB 421 CTATGGGCAAGGCAAGCACTGCACTGCACTCAAGACCGGAGGATGAAACAAAGCCAC 480  
QY 481 TGAAGCTAGTCAACCTCAATCTAGCAAGGGGGTTACAAACAGCCAGCTTATGATATG 540  
DB 481 TGAAGCTAGTCAACCTCAATCTAGCAAGGGGGTTACAAACAGCCAGCTTATGATATG 540  
QY 541 ACAAGATTAATCAAGTATATCCCAAGGTTACTGGGAGCTACCCCAATGACAGCACTGC 600  
DB 541 ACAAGATTAATCAAGTATATCCCAAGGTTACTGGGAGCTACCCCAATGACAGCACTGC 600  
QY 601 ACCCTCACTCTCAACCTCTCAAGCTATTCCTCTACACAGCCGACTGATATGATCAAG 660  
DB 601 ACCCTCACTCTCAACCTCTCAAGCTATTCCTCTACACAGCCGACTGATATGATCAAG 660  
QY 661 CAGTACTCTCAGAGAACTCTATGGGCAACCGAGAGCTATGACAGAGAGTACTA 720  
DB 661 CAGTACTCTCAGAGAACTCTATGGGCAACCGAGAGCTATGACAGAGAGTACTA 720  
QY 721 TGGTCAACAAAGAGCTATGGGCAAGGCTCCCACTGATTAACCACTCCCACTGATC 780  
DB 721 TGGTCAACAAAGAGCTATGGGCAAGGCTCCCACTGATTAACCACTCCCACTGATC 780  
QY 781 CTACAGCAAGCTCCAGTCAATATAGCAACAGAGAGGCTACGGGAGAGAGTTC 840  
DB 781 CTACAGCAAGCTCCAGTCAATATAGCAACAGAGAGGCTACGGGAGAGAGTTC 840  
QY 841 ATTCCGACAGAGCAACCCAGTATGAGTGGTGTATATGGGCAAGAGTCTGAGGATTTTC 900  
DB 841 ATTCCGACAGAGCAACCCAGTATGAGTGGTGTATATGGGCAAGAGTCTGAGGATTTTC 900  
QY 901 CGGACCAAGAGAACTGGAGCATGAGTGGCTCTGATTAACCGGGGAGAGGAGGAG 960  
DB 901 CGGACCAAGAGAACTGGAGCATGAGTGGCTCTGATTAACCGGGGAGAGGAGGAG 960  
QY 961 ATTGATCGTGGAGGCAATGAGAGAGGTCGGGGGAGAGGAGCGCGGTGGAATGGGAG 1020  
DB 961 ATTGATCGTGGAGGCAATGAGAGAGGTCGGGGGAGAGGAGCGCGGTGGAATGGGAG 1020  
QY 1021 CGCTGGAGAGCGAGTGGCTTCAATAGCTGTGTGAGCCCATGGATGAAAGCAAGTCT 1080  
DB 1021 CGCTGGAGAGCGAGTGGCTTCAATAGCTGTGTGAGCCCATGGATGAAAGCAAGTCT 1080  
QY 1081 TGAATCAAGGCGCACTGTATATCAAGTGAAGCTGCAACAGTGCATTTATATGACA 1140  
DB 1081 TGAATCAAGGCGCACTGTATATCAAGTGAAGCTGCAACAGTGCATTTATATGACA 1140  
QY 1141 AGGATTAATGAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 1200  
DB 1141 AGGATTAATGAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 1200  
QY 1201 TGTAAAGATGAG 1260  
DB 1201 TGTAAAGATGAG 1260

QY 1261 AGGAAAGCCCAAGAGGAGATGCCAGTGTCTTATGAAGACCACTGCTCCAGAGCTGC 1320  
DB 1261 AGGAAAGCCCAAGAGGAGATGCCAGTGTCTTATGAAGACCACTGCTCCAGAGCTGC 1320  
QY 1321 CGTGAATGGTGTGATGGGAAAGATTTTCAAGGAGAGCAAACTTAAAGTCTCCCTGCTGC 1380  
DB 1321 CGTGAATGGTGTGATGGGAAAGATTTTCAAGGAGAGCAAACTTAAAGTCTCCCTGCTGC 1380  
QY 1381 GAAAGAGCTCTCAATGAACATATGCGGGGTGTCTGCCACCCCTGAGAGGAGAGCAT 1440  
DB 1381 GAAAGAGCTCTCAATGAACATATGCGGGGTGTCTGCCACCCCTGAGAGGAGAGCAT 1440  
QY 1441 GCCAACCACTCTCGTGAAGTGTCCAGAGGCTCCAGAGAGTCTCGGGGAGACCTATGGGTGC 1500  
DB 1441 GCCAACCACTCTCGTGAAGTGTCCAGAGGCTCCAGAGAGTCTCGGGGAGACCTATGGGTGC 1500  
QY 1501 CATGGAGGCGGTGAGAGAGATGAGAGAGGCTTCCCTCCAAAGAGAGACCCCGGGGTTCCCG 1560  
DB 1501 CATGGAGGCGGTGAGAGAGATGAGAGAGGCTTCCCTCCAAAGAGAGACCCCGGGGTTCCCG 1560  
QY 1561 AGGAAACCCCTCTGAGAGAGAGAAAGCTCCAGACCGAGCTGAGACTGGCATGTCCCA 1620  
DB 1561 AGGAAACCCCTCTGAGAGAGAGAAAGCTCCAGACCGAGCTGAGACTGGCATGTCCCA 1620  
QY 1621 TCCGGGTGTGAGAAACCAAGACTTCCCTGAGAGACAGAGTGCACCAAGTGAAGCCCC 1680  
DB 1621 TCCGGGTGTGAGAAACCAAGACTTCCCTGAGAGACAGAGTGCACCAAGTGAAGCCCC 1680  
QY 1681 AAGAGCTGAAGGCTTCTCCCGGCAACCTTCCCGGCGGGGAGTGTGAGGAGAGG 1740  
DB 1681 AAGAGCTGAAGGCTTCTCCCGGCAACCTTCCCGGCGGGGAGTGTGAGGAGAGG 1740  
QY 1741 TGGCCCTGTGGCATCGGGAGAGAAAGAGTGGCTCATGATGTGTGTCCCGGTGG 1800  
DB 1741 TGGCCCTGTGGCATCGGGAGAGAAAGAGTGGCTCATGATGTGTGTCCCGGTGG 1800  
QY 1801 AATGTTCAAGAGTGGCTGTGTGTGAGACAGAGTGGCTTCCGTGTGTGTGTGTGTGTGT 1860  
DB 1801 AATGTTCAAGAGTGGCTGTGTGTGAGACAGAGTGGCTTCCGTGTGTGTGTGTGTGTGT 1860  
QY 1861 CCAGAGTGGCTTGTGTGTGAGAGAAAGAGAGTGGCTTCCGTGTGTGTGTGTGTGTGT 1920  
DB 1861 CCAGAGTGGCTTGTGTGTGAGAGAAAGAGAGTGGCTTCCGTGTGTGTGTGTGTGTGT 1920  
QY 1921 GGAACAGATGGAGAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1980  
DB 1921 GGAACAGATGGAGAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1980  
QY 1981 GCAACGTCAG 2040  
DB 1981 GCAACGTCAG 2040  
QY 2041 GACTACAGATTTATTTTAAACAGAAAGTGTAAATTAATTAATTAATTAATTAATTAAT 2100  
DB 2041 GACTACAGATTTATTTTAAACAGAAAGTGTAAATTAATTAATTAATTAATTAATTAAT 2100  
QY 2101 ATGTTGGCAACAATTAATGATTAATCTTGTCTGTATAGTATTTTCAACATTTGT 2160  
DB 2101 ATGTTGGCAACAATTAATGATTAATCTTGTCTGTATAGTATTTTCAACATTTGT 2160  
QY 2161 GAAAGAACTTTAAACAAAGTAAATGTATGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2220  
DB 2161 GAAAGAACTTTAAACAAAGTAAATGTATGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2220  
QY 2221 AAAAATGT 2280  
DB 2221 AAAAATGT 2280  
QY 2281 GTGAGAACCAAGAGGCTCTTAACTGTAACTGTAACTGTAACTGTAACTGTAACTGTAACT 2340  
DB 2281 GTGAGAACCAAGAGGCTCTTAACTGTAACTGTAACTGTAACTGTAACTGTAACTGTAACT 2340  
QY 2341 TTTTAAATTAATTAATTCCAATGTGTAAATTAATTAATTAATTAATTAATTAATTAAT 2390















Db 728 GATTCCTACAGCAAGCTCCAGTCAATATAGCCAAACAGACGACGCTACGGGACGACG 787  
 Qy 836 AGTTCAATCCGACAGACCAACCCCACTAGACATGGGTGTTTATGGGACGAGTCTGGAGCA 895  
 Db 788 AGTTCAATCCGACAGACCAACCCCACTAGACATGGGTGTTTATGGGACGAGTCTGGAGCA 847  
 Qy 896 TTTTCGGACACAGAGAGAACCGGAGCATGATGGGCTGATTAACCGGGGACAGGGGAGA 955  
 Db 848 TTTTCGGACACAGAGAGAACCGGAGCATGATGGGCTGATTAACCGGGGACAGGGGAGA 907  
 Qy 956 GGGGGAATTTGATCTGGAGGACATGACAGAGTGGGCGGGAGAGAGACCGGTGGATG 1015  
 Db 908 GGGGGAATTTGATCTGGAGGACATGACAGAGTGGGCGGGAGAGAGACCGGTGGATG 967  
 Qy 1016 GGCAGGCTGAGAGAGAGAGTGGCTTCAATTAACCTGTGTGACCCATGATGAAGACCA 1075  
 Db 968 GGCAGGCTGAGAGAGAGAGTGGCTTCAATTAACCTGTGTGACCCATGATGAAGACCA 1027  
 Qy 1076 GATCTTGAATCTAGACCCACCTGTAGATCCAGATGAAGACCTGACAAACAGTGCATTTAT 1135  
 Db 1028 GATCTTGAATCTAGACCCACCTGTAGATCCAGATGAAGACCTGACAAACAGTGCATTTAT 1087  
 Qy 1136 GTAACAAGATTAATATGACAGTGTGATCTAGATGATCTGGACAGCTTCTTTAAGCAGTGT 1195  
 Db 1088 GTACAAAGATTAATATGACAGTGTGATCTAGATGATCTGGACAGCTTCTTTAAGCAGTGT 1147  
 Qy 1196 GGGGTTTAAATGATGAACAAGAGAACTGGGCAACCAATGATCAATCTAAGCTGGACAG 1255  
 Db 1148 GGGGTTTAAATGATGAACAAGAGAACTGGGCAACCAATGATCAATCTAAGCTGGACAG 1207  
 Qy 1256 GAAACAGAGAAAGCCCAAGGGGATGCAAGTGTCTTAAAGAAACCAACCACTGACAG 1315  
 Db 1208 GAAACAGAGAAAGCCCAAGGGGATGCAAGTGTCTTAAAGAAACCAACCACTGACAG 1267  
 Qy 1316 GCTGCGGTGGAATGTTTGAAGGAAAGATTTTCAAGGAGCAACTTAAGTCTCCCTT 1375  
 Db 1268 GCTGCGGTGGAATGTTTGAAGGAAAGATTTTCAAGGAGCAACTTAAGTCTCCCTT 1327  
 Qy 1376 GCTTCGGAAGAGCTTCAATGAACAATGTCGGGGTGTGTGTCACCCCGTGAAGGCAAG 1435  
 Db 1328 GCTTCGGAAGAGCTTCAATGAACAATGTCGGGGTGTGTGTCACCCCGTGAAGGCAAG 1387  
 Qy 1436 GGCATGCAACCACTCCGTGAGAGTCCAGAGGCGCCAGAGAGCTCTGGGGAGACCAAG 1495  
 Db 1388 GGCATGCAACCACTCCGTGAGAGTCCAGAGGCGCCAGAGAGCTCTGGGGAGACCAAG 1447  
 Qy 1496 GGTGCGATGGAGGCGGTGAGAGATAGAGAGGCTTCCCTTCAAGAGACCCCGGGGT 1555  
 Db 1448 GGTGCGATGGAGGCGGTGAGAGATAGAGAGGCTTCCCTTCAAGAGACCCCGGGGT 1507  
 Qy 1556 TCCCGAGGGAACCCCTCTGAGAGAGAAAGTTCACGACCGAGCTGAGACCTGAGTGT 1615  
 Db 1508 TCCCGAGGGAACCCCTCTGAGAGAGAAAGTTCACGACCGAGCTGAGACCTGAGTGT 1567  
 Qy 1616 CCCAAATCCCGGTTGTGAGAAACAGAACTTCCCTGAGAGAAACAGAGTCAACCAAGTGTAG 1675  
 Db 1568 CCCAAATCCCGGTTGTGAGAAACAGAACTTCCCTGAGAGAAACAGAGTCAACCAAGTGTAG 1627  
 Qy 1676 GCCCCAAAGCTGAGAGGCTTCTCCCGCAACCTTTCGCGCCCGGGGTGTGATCTGTGC 1735  
 Db 1628 GCCCCAAAGCTGAGAGGCTTCTCTCCCGCAACCTTTCGCGCCCGGGGTGTGATCTGTGC 1687  
 Qy 1736 AAGAGTGGCCCTGTGTGACATGCGGGAGAGAAAGTGGCTTCAATGATCTGTGTGTCC 1795  
 Db 1688 AAGAGTGGCCCTGTGTGACATGCGGGAGAGAAAGTGGCTTCAATGATCTGTGTGTCC 1747  
 Qy 1796 GGTGGAATGTTCAAGAGTGGCGGTGAGAGACAGAGTGTGCTTCCGTGTGTGTCCGGGG 1854  
 Db 1748 GGTGGAATGTTCAAGAGTGGCGGTGAGAGACAGAGTGTGCTTCCGTGTGTGTCCGGGG 1807  
 Qy 1855 CATGACCGAGGTGTGTTGTGTGTGAGAGAAACAGAGTGTGCTTGTGGGGGCCCCCTGAGAC 1914  
 Db 1808 CATGACCGAGGTGTGTTGTGTGTGAGAGAAACAGAGTGTGCTTGTGGGGGCCCCCTGAGAC 1867

Qy 1915 TTTGATGAAACAGATGGAGAGAAAGAGAGAGAGAGCTGAGAGACCTGGAATAATGATTA 1974  
 Db 1868 TTTGATGAAACAGATGGAGAGAAAGAGAGAGAGAGCTGAGAGACCTGGAATAATGATTA 1927  
 Qy 1975 AGGCGAGACCCGTCAGAGCGGAGAGATGGGCTCTACTAGTACAGAGACCCCGAGAGC 2034  
 Db 1928 AGGCGAGACCCGTCAGAGCGGAGAGATGGGCTCTACTAGTACAGAGACCCCGAGAGC 1987  
 Qy 2035 TGCATTGACTACAGATTTATTTTAAACAGAAAATGTTTAAATTTAATTCATTA 2094  
 Db 1988 TGCATTGACTACAGATTTATTTTAAACAGAAAATGTTTAAATTTAATTCATTA 2047  
 Qy 2095 TTTAATATGTTGGCCCAAC-ATTATGATTTATCTTGTCTGACTTTAGATTTTTCAC 2153  
 Db 2048 TTTAATATGTTGGCCCAACAAATTAATGATTTATCTTGTCTGACTTTAGATTTTTCAC 2107  
 Qy 2154 CATTTGTAAGAAACATTAATAAAGTAAATGATGAGTGGGAGATTTTTCCTCC 2213  
 Db 2108 CATTTGTAAGAAACATTAATAAAGTAAATGATGAGTGGGAGATTTTTCCTCC 2167  
 Qy 2214 TTTCTTTAAAAATGTTGTTTAAAGCTTTAACAATGGAACCCCTGTGAGCATGCTCAG 2273  
 Db 2168 TTTCTTTAAAAATGTTGTTTAAAGCTTTAACAATGGAACCCCTGTGAGCATGCTCAG 2227  
 Qy 2274 TATCATTTGAGAGAACCAAGAGGCTCTTAACCTGTAACAATGTTGATGATGTTT 2333  
 Db 2228 TATCATTTGAGAGAACCAAGAGGCTCTTAACCTGTAACAATGTTGATGATGTTT 2287  
 Qy 2334 TTTTCTTTTAAATTAATAATTCATAATGTTAATTA 2373  
 Db 2288 TTTTCTTTTAAATTAATAATTCATAATGTTAATA 2326

RESULT 8  
 CR456490  
 LOCUS  
 DEFINITION  
 Homo sapiens EWSR1 full length open reading frame (ORF) cDNA clone  
 (cDNA clone C220RF:pgEM.EWSR1).  
 ACCESSION  
 CR456490.1 GI:47678510  
 VERSION  
 CDNA: chromosome 22: ORF.  
 KEYWORDS  
 Homo sapiens (human)  
 SOURCE  
 ORGANISM  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE  
 Collins, J.E., Wright, C.L., Edwards, C.A., Davis, M.P., Grinham, J.A.,  
 1 (bases 1 to 2189)  
 Huckle, E.J., Beate, D.M. and Dunham, I.  
 DIRECT SUBMISSION  
 Submitted (24-MAY-2004) Sanger Institute, Hinxton, Cambridgeshire,  
 CB10 1SA, UK. E-mail enquiries: c220@sanger.ac.uk Manuscript  
 Sanger Institute name: pgEM.EWSR1  
 COMMENT  
 Homo sapiens CDNA sequence. This sequence was generated as part of  
 The Wellcome Trust Sanger Institute program to isolate cDNA clones  
 representing the full length open reading frame of well annotated  
 protein coding genes on human chromosome 22. For more information  
 see <http://www.sanger.ac.uk/HGP/Chr22/>.  
 FEATURES  
 source  
 1..2189  
 location/Qualifiers  
 1..2189  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /chromosome="22"  
 /clone="pgEM.EWSR1"  
 /lab\_host="JM109"  
 1..2189  
 /gene="EWSR1"  
 34..2001  
 /gene="EWSR1"  
 /codon\_start=1  
 /protein\_id="CAG30376.1"



DB	Accession	Version	Keywords	Source	Organism	Reference	Authors
Oy	1931	CGAGGAAAGAGGAGGACGTGAGAGACCTGGAAAAATGGATTAAGGAGACACCGTAG	1990				
Db	1918	CGAGGAAAGAGGAGGACGTGAGAGACCTGGAAAAATGGATTAAGGAGACACCGTAG	1977				
Oy	1991	GAGCGAGAGATCGGCGCTACTAGATGACAGAGACCCGCGAGAGCTGCATTTAGTACACGA	2050				
Db	1978	GAGCGCGAGATCGGCGCTACTAGATGACAGAGACCCCGAGAGCTGCATTTAGTACACGA	2037				
Oy	2051	TTTATTTTAAACAGAAAATGTTTAAATTTAATTCATATTTAATTTGTTGGCA	2110				
Db	2038	TTTATTTTAAACAGAAAATGTTTAAATTTAATTCATATTTAATTTGTTGGCA	2097				
Oy	2111	CAACATTTGATTTATTCCTGCTGCTACTTATTTAGTATTTTTCACCATTTGTGAGAAACAT	2170				
Db	2098	CAACATTTGATTTATTCCTGCTGCTACTTATTTAGTATTTTTCACCATTTGTGAGAAACAT	2157				
Oy	2171	TAAACAGATTAAATGTAGTGTGCGAGATT	2202				
Db	2158	TAAACAGATTAAATGTAGTGTGCGAGATT	2189				
RESULT 9	LOCUS	BC004817					
DEFINITION	BC004817	2182 bp	mRNA	linear	PRI 29-JUN-2004		
ACCESSION	BC004817						
VERSION	BC004817.1	GI:13435962					
KEYWORDS	MG.						
SOURCE	MG.						
ORGANISM	MG.						
REFERENCE	MG.						
AUTHORS	MG.						
ORGANISM	MG.						
DEFINITION	MG.						
ACCESSION	MG.						
VERSION	MG.						
KEYWORDS	MG.						
SOURCE	MG.						
ORGANISM	MG.						
DEFINITION	MG.						
ACCESSION	MG.						
VERSION	MG.						
KEYWORDS	MG.						
SOURCE	MG.						
ORGANISM	MG.						
DEFINITION	MG.						
ACCESSION	MG.						
VERSION	MG.						
KEYWORDS	MG.						
SOURCE	MG.						
ORGANISM	MG.						
DEFINITION	MG.						
ACCESSION	MG.						
VERSION	MG.						
KEYWORDS	MG.						
SOURCE	MG.						
ORGANISM	MG.						
DEFINITION	MG.						
ACCESSION	MG.						
VERSION	MG.						
KEYWORDS	MG.						
SOURCE	MG.						
ORGANISM	MG.						
DEFINITION	MG.						
ACCESSION	MG.						
VERSION	MG.						
KEYWORDS	MG.						
SOURCE	MG.						
ORGANISM	MG.						
DEFINITION	MG.						
ACCESSION	MG.						
VERSION	MG.						
KEYWORDS	MG.						
SOURCE	MG.						
ORGANISM	MG.						
DEFINITION	MG.						
ACCESSION	MG.						
VERSION	MG.						
KEYWORDS	MG.						
SOURCE	MG.						
ORGANISM	MG.						
DEFINITION	MG.						

Web site: <http://www.hpsc.bcm.tmc.edu/cdna/>  
 Contact: amg@bcm.tmc.edu  
 Kowars, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Louised, H.,  
 Kowars, C.R., Sneed, A.J., Martin, R.G., Muzny, D.M., Nannavati,  
 A.N., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found  
 through the I.M.A.G.E. Consortium/Link at: <http://image.lnl.gov>  
 Series: IRAX Plate: 3 Row: 1 Column: 6  
 This clone was selected for full length sequencing because it  
 passed the following selection criteria: Hexamer frequency ORF  
 analysis.

location/Qualifiers

```

1. 2182
  /organism="Homo sapiens"
  /mol_type="mRNA"
  /db_xref="taxon:9606"
  /clone="MGC:5432 IMAGE:349145"
  /issue_type="Placenta, chorioecarcinoma"
  /clone_lib="NIH MGC_10"
  /lab_host="DH10B"
  /note="Vector: pCMV-SPORT6"

1. 2182
  /gene="EWSR1"
  /note="synonym: EMS"
  /db_xref="LocustID:2130"
  /db_xref="MIM:133450"
  20. 1890
  /gene="EWSR1"
  /codon_start=1
  /product="Ewing sarcoma breakpoint region 1, isoform EMS"
  /protein_id="AA04817.1"
  /db_xref="GI:1345963"
  /db_xref="LocustID:2130"
  /db_xref="MIM:133450"

  /translation="MASTDYSTYSQAQAQGSAYTAQTPQGYAQTTOAYGQGSYGYT
  GQPDVYTOAQATYAGTAVATSYGQPTATTPAPAYSGPVGYSGTGADDTT
  AATVTTQASVQAASAVGTOPAPVAYGGQAPAAATPTRPDGNKPLETSPSGVNO
  PSIGYSGSYSPDVGSSYPMOPVTPAPSPYPTSPSTSYDQSSYQOONYGQPS
  SYGQGSYTGQGSSTGQOPTSYPTSPQTSYQAPSYQSSYTGQSSYFQDHPSSNG
  VYQGSYGFSFGPEGNRMSGPDMRGKRGFDRGMSRGSGRGSGRGSAGRGGSNG
  KPGPDMEGPDLDLGPVDEDESDNSAIYVQGLNDSVTLDDLDLDFKQCGVAKMR
  TGQPMIHIYLDKETKGDAVSYEDPYPAKAAVEVDFDKDFQGSGLKSLAKRKP
  MNSRRGGLPREGGMPPRLRGFGGFGGGMGRGGRGGRGGRGGRGGRGGRGGR
  PSGGGAYQHRAGDMQCCGNQNPMPATRECMQCTAPKEGTLPPPPPGGDRGSG
  GPGMRGGRGGLMDRGPGCMFRGGRGSDGGRTRGGRGMDRGGRGGRGGRGGRGGR
  LMEQGRGRGGRGGRGGRMDRGGRHQRDRPPY"

```

ORIGIN

```

Query Match      90.6%; Score 2165; DB 9; Length 2182;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2165; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

Qy	25	CGAGAGGAAAGAGAGAAAAATGGCGTCACAGATTACGATCCTATAGCCAAAGCTGAGC	84
Db	1	CGAGAGGAAAGAGAGAAAAATGGCGTCACAGATTACGATCCTATAGCCAAAGCTGAGC	60
Qy	85	GCAGCAGGCGTACAGTCTTACACCGCCAGCCCACTCAAGATATGCAAGACCA	144
Db	61	GCAGCAGGCGTACAGTCTTACACCGCCAGCCCACTCAAGATATGCAAGACCA	120
Qy	145	GGCATATGGGCAACAAAGCTATGGAACTTAATGACAGCCCACTGATGTCAGCTATACCA	204
Db	121	GGCATATGGGCAACAAAGCTATGGAACTTAATGACAGCCCACTGATGTCAGCTATACCA	180
Qy	205	GGCTCAGACCACTGCAACTATGGGAGACCGCCTATGCAACTTCTTATGACAGCCTCC	264
Db	181	GGCTCAGACCACTGCAACTATGGGAGACCGCCTATGCAACTTCTTATGACAGCCTCC	240
Qy	265	CACGTGTTATCTACTCACTGCCCCCGCCAGGATACAGCCACCTGTCCAGGGGATATGG	324
Db	241	CACGTGTTATCTACTCACTGCCCCCGCCAGGATACAGCCACCTGTTCAGGGGATATGG	300



QY 325 CACTGGTGTATGATATACCACTGCTACAGTCAACCAACCCAGGCTCTATGACG 384  
 DB 301 CACTGGTGTATGATATACCACTGCTACAGTCAACCAACCCAGGCTCTATGACG 360  
 QY 385 TCAGTGTGATATGATGACCTGACCTGCTATTCAGGCTATGAGGACAGCAGCAGC 444  
 DB 361 TCAGTGTGATATGATGACCTGACCTGCTATTCAGGCTATGAGGACAGCAGCAGC 420  
 QY 445 TGCAGCTACAAAGACCCAGGATGAAACAAAGCCCACTGAGCACTAGTCAACTCAATCTAG 504  
 DB 421 TGCAGCTACAAAGACCCAGGATGAAACAAAGCCCACTGAGCACTAGTCAACTCAATCTAG 480  
 QY 505 CACAGGAGGTTTAAACCAAGCCAGCTAGGATATGACAGAGTATACAGTATATCCCA 564  
 DB 481 CACAGGAGGTTTAAACCAAGCCAGCTAGGATATGACAGAGTATATCCCA 540  
 QY 565 GGTACTGGAGGTATCCCACTGAGCAGTCACTGACCTCAATCCCTCTACAG 624  
 DB 541 GGTACTGGAGGTATCCCACTGAGCAGTCACTGACCTCAATCCCTCTACAG 600  
 QY 625 CTATTCCTCTACACAGCCCACTAGTATATGATCAGAGCACTATCTCAGAGAAACCTTA 684  
 DB 601 CTATTCCTCTACACAGCCCACTAGTATATGATCAGAGCACTATCTCAGAGAAACCTTA 660  
 QY 685 TGGGCAACCGAGAGCTATGACAGAGAGTATGATGATCAAAAGCAGTATGAGCA 744  
 DB 661 TGGGCAACCGAGAGCTATGACAGAGAGTATGATGATCAAAAGCAGTATGAGCA 720  
 QY 745 GCAAGCTCCCACTAGTATCCCACTGAGCACTGATCTCAAGCTCAAGTCAAT 804  
 DB 721 GCAAGCTCCCACTAGTATCCCACTGAGCACTGATCTCAAGCTCAAGTCAAT 780  
 QY 805 TACCCAAACAGAGAGAGTATGAGGAGAGTATGATGATCAGAGCACTCAAGCTCAAGT 864  
 DB 781 TACCCAAACAGAGAGAGTATGAGGAGAGTATGATGATCAGAGCACTCAAGCTCAAGT 840  
 QY 865 CATGGGTTTATGAGGAGAGTCTGAGAGATTTTCGAGCAAGAGAGAGAGAGTATGAG 924  
 DB 841 CATGGGTTTATGAGGAGAGTCTGAGAGATTTTCGAGCAAGAGAGAGAGAGTATGAG 900  
 QY 925 GAGTGGCTCTGATTAACCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 984  
 DB 901 GAGTGGCTCTGATTAACCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 960  
 QY 985 AGGTGGGCGGAG 1044  
 DB 961 AGGTGGGCGGAG 1020  
 QY 1045 TAAAGCTGTGAGCCATGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1104  
 DB 1021 TAAAGCTGTGAGCCATGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1080  
 QY 1105 AGATGAAGTCTGACCAACAGTGAATTTATGATCAAGATTTAATGACAGTGTACTCT 1164  
 DB 1081 AGATGAAGTCTGACCAACAGTGAATTTATGATCAAGATTTAATGACAGTGTACTCT 1140  
 QY 1165 AGATGAAGTCTGACCAACAGTGAATTTATGATCAAGATTTAATGACAGTGTACTCT 1224  
 DB 1141 AGATGAAGTCTGACCAACAGTGAATTTATGATCAAGATTTAATGACAGTGTACTCT 1200  
 QY 1225 GCAACCATGATCAATCTACCTGACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1284  
 DB 1201 GCAACCATGATCAATCTACCTGACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1260  
 QY 1285 AGTGTCTATGAG 1344  
 DB 1261 AGTGTCTATGAG 1320  
 QY 1345 TTTTCAAGGAG 1404  
 DB 1321 TTTTCAAGGAG 1380  
 QY 1405 GCGGGGTGTCTGCAACCCGCTGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1464

DB 1381 GCGGGGTGTCTGCAACCCGCTGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1440  
 QY 1465 AGAGAGCCCAAGAGAGTCTGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1524  
 DB 1441 AGAGAGCCCAAGAGAGTCTGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1500  
 QY 1525 AGAGAGCTTCCCTCAAG 1584  
 DB 1501 AGAGAGCTTCCCTCAAG 1560  
 QY 1585 CGTCCAGACCGAGCTGAG 1644  
 DB 1561 CGTCCAGACCGAGCTGAG 1620  
 QY 1645 CGCCTGAG 1704  
 DB 1621 CGCCTGAG 1680  
 QY 1705 ACCCTTCCGCCCCGAGGAG 1764  
 DB 1681 ACCCTTCCGCCCCGAGGAG 1740  
 QY 1765 AAGAGGTGCTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1824  
 DB 1741 AAGAGGTGCTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1800  
 QY 1825 AAGAGAGGTGCTTCCGAG 1884  
 DB 1801 AAGAGAGGTGCTTCCGAG 1860  
 QY 1885 ACGAGGTGCTTCCGAG 1944  
 DB 1861 ACGAGGTGCTTCCGAG 1920  
 QY 1945 ACGAGGTGAG 2004  
 DB 1921 ACGAGGTGAG 1980  
 QY 2005 GGCCTATGATGAG 2064  
 DB 1981 GGCCTATGATGAG 2040  
 QY 2065 CAGAAATGTTTAAATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTA 2124  
 DB 2041 CAGAAATGTTTAAATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTA 2100  
 QY 2125 TTCCTGTCTGATCTTATTTTACCATTTTGTGAAGAGAGAGAGAGAGAGAGAGAGAG 2184  
 DB 2101 TTCCTGTCTGATCTTATTTTACCATTTTGTGAAGAGAGAGAGAGAGAGAGAGAGAG 2160  
 QY 2185 TGGTA 2189  
 DB 2161 TGGTA 2165

RESULT 10  
 BC072442 2164 bp mRNA linear PRI 30-JUN-2004  
 LOCUS Homo sapiens Ewing sarcoma breakpoint region 1, transcript variant  
 DEFINITION EMS, mRNA (CDNA clone MGC:87992 IMAGE:6047805), complete cds.  
 ACCESSION BC072442  
 VERSION BC072442.1 GI:48734726  
 KEYWORDS MGC.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 2164)  
 AUTHORS Strausberg, R.L., Feingold, E.A., Grouse, L.H., Dege, J.G.,  
 Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D.,  
 Altshuler, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K.,  
 Hopkins, R.F., Jordan, H., Moore, T., Wang, S.I., Wang, J., Heide, F.,



Datchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L., Shepleton, M., Soares, M.B., Bonaldi, M.P., Casavant, T.L., Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Rosenthal, S., Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J., Abramson, R.D., Mullaly, S.J., Bosak, S.A., McEwan, P.J., McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S., Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W., Villalon, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Fahey, J., Helton, E., Kettelman, M., Madan, A., Rodriguez, S., Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y., Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D., Dickinson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Bickford, Y.S., Krzywinski, M.I., Skalska, U., Smalins, D.E., Schermer, A., Schein, J.E., Jones, S.J. and Marra, M.A. Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences  
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)

12477932

2 (bases 1 to 2164)

Strausberg, R.  
Direct Submission  
Submitted (01-JUN-2004) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

NIH-MGC Project URL: <http://mgc.ncl.nih.gov>

REMARK  
COMMENT  
Contact: MGC help desk  
Email: [cgaps-rc@mail.nih.gov](mailto:cgaps-rc@mail.nih.gov)  
Tissue Procurement: ATCC/DCTD/DRP  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)  
DNA Sequencing by: National Institutes of Health Intramural Sequencing Center (NISC),  
Gaithersburg, Maryland;  
Web site: <http://www.nisc.nih.gov/>  
Contact: [nisc\\_mgc@mail.nih.gov](mailto:nisc_mgc@mail.nih.gov)  
Akher, N., Ayele, K., Beckstrom-Sternberg, S.M., Benjamin, B., Blakesley, R.W., Bouffard, G.G., Breen, K., Brinkley, C., Brooks, S., Dietrich, N.L., Granite, S., Guan, X., Gupta, J., Haghighi, P., Hansen, N., Ho, S.-L., Karling, E., Kwong, P., Laric, P., Legaspi, R., Maduro, Q.L., Mastello, C., Maskeri, B., Mastrian, S.D., McCloskey, J.C., McDowell, J., Pearson, R., Stantirpop, S., Thomas, P.J., Touchman, J.W., Tauson, C., Vogt, J.L., Walker, M.A., Metherby, K.D., Wiggins, L., Young, A., Zhang, L.-H. and Green, E.D.

FEATURES  
source  
Location/Qualifiers  
1..2164  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="MGC:87992 IMAGE:6047805"  
/issue\_type="Skin, melanotic melanoma."  
/clone\_1ib="NIH\_MGC\_72"  
/lab\_host="DH10B"  
/note="Vector: pCMV-SPORT6"  
1..2164  
/gene="EWSR1"  
/note="synonym: EWS"  
/db\_xref="LocustID:2130"  
/db\_xref="MIM:133450"  
7..1974  
/gene="EWSR1"  
/codon\_start=1  
/product="Ewing sarcoma breakpoint region 1, isoform EWS"  
/protein\_id="AAH72422.1"  
/db\_xref="GI:48734727"  
/db\_xref="LocustID:2130"  
/db\_xref="MIM:133450"

## ORIGIN

Query Match 89.4%; Score 2136; DB 9; Length 2164;  
Best Local Similarity 99.9%; Pred. No. 0;  
Matches 2149; Conservative 0; Mismatches 0; Indels 3; Gaps 1;

/translation="MASTDYSTYSOAAAGQVGAATPQVGAQTQTAAYGOOSYRT  
GQPTDVSYTOAQTATATYGTATATSTGQPTGTTTPAPATQVQVGTGAVDTT  
ATVTTQASTAASQASVAGTQAPVAVYQAPATAPTPQDNKRTESQPSSTVGNO  
PSLGGQSSYQVQVQVPSYPMQVAVYQAPATAPTPQDNKRTESQPSSTVGNO  
SYGOOSYGOOSYGOOPPSYPPQRTGSYQVQVQVQVQVQVQVQVQVQVQVQV  
VYGOOSYGOOSYGOOSYGOOSYGOOSYGOOSYGOOSYGOOSYGOOSYGOOS  
PGPMRGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGG  
GQPMHITLYDKETGRKGDATYSYEDPTAKAVBNFDGDFGSKLKSLAKRPP  
NSMRKLPREGRGMPPLRGQVGGGPGGPGGPGGPGGPGGPGGPGGPGGPGG  
SGGANYQVHAGMDQCPNPGQGNQNFMRTECNQCKAPKEGFLPPPPPGGDRGSG  
PGMRGSGGGLMDRGSGGPGGPGGPGGPGGPGGPGGPGGPGGPGGPGGPGG  
MEQMGRRGSGGSGGPGGMDGEGHQRDRNPY"

38 GAGAAATGCGCGCCACGATTTACATTAAGCCAGAGCGGACGAGGCTAC 97  
1 GAGAAATGCGCGCCACGATTTACATTAAGCCAGAGCGGACGAGGCTAC 60  
98 AGTCTTACACGCGCCACGATTTACATTAAGCCAGAGCGGACGAGGCTAC 157  
61 AGTCTTACACGCGCCACGATTTACATTAAGCCAGAGCGGACGAGGCTAC 120  
158 CAAAGCTATGAACTATGAGCAGCCCACTGATGTCAGTATACCAAGCTCAGACCACT 217  
121 CAAAGCTATGAACTATGAGCAGCCCACTGATGTCAGTATACCAAGCTCAGACCACT 180  
218 GCAACCTATGAGGAGACCGCTATGAGCAGCCCACTGATGTCAGTATACCAAGCTCAGACCACT 277  
181 GCAACCTATGAGGAGACCGCTATGAGCAGCCCACTGATGTCAGTATACCAAGCTCAGACCACT 240  
278 ACTCAACTGCGCGCCACGATTTACATTAAGCCAGAGCGGACGAGGCTAC 337  
241 ACTCAACTGCGCGCCACGATTTACATTAAGCCAGAGCGGACGAGGCTAC 300  
338 GATACCACTGCTTACATGACGACCACTGATGTCAGTATACCAAGCTCAGACCACT 397  
301 GATACCACTGCTTACATGACGACCACTGATGTCAGTATACCAAGCTCAGACCACT 360  
398 GGAATCACTGCTTACATGACGACCACTGATGTCAGTATACCAAGCTCAGACCACT 457  
361 GGAATCACTGCTTACATGACGACCACTGATGTCAGTATACCAAGCTCAGACCACT 420  
458 CGCAGATGAGAAACAGCCCACTGATGTCAGTATACCAAGCTCAGACCACT 517  
421 CGCAGATGAGAAACAGCCCACTGATGTCAGTATACCAAGCTCAGACCACT 480  
518 AACCAAGCCAGCTATGAGCAGCCCACTGATGTCAGTATACCAAGCTCAGACCACT 577  
481 AACCAAGCCAGCTATGAGCAGCCCACTGATGTCAGTATACCAAGCTCAGACCACT 540  
578 TACCCATGAGCAGCTATGAGCAGCCCACTGATGTCAGTATACCAAGCTCAGACCACT 637  
541 TACCCATGAGCAGCTATGAGCAGCCCACTGATGTCAGTATACCAAGCTCAGACCACT 600  
638 CAGCCAGCTATGAGCAGCCCACTGATGTCAGTATACCAAGCTCAGACCACT 697  
601 CAGCCAGCTATGAGCAGCCCACTGATGTCAGTATACCAAGCTCAGACCACT 660  
698 AGCTATGAGCAGCAGTATGAGCAGCCCACTGATGTCAGTATACCAAGCTCAGACCACT 757  
661 AGCTATGAGCAGCAGTATGAGCAGCCCACTGATGTCAGTATACCAAGCTCAGACCACT 720  
758 AGTTACCCACCCCACTGATGTCAGCAGCCCACTGATGTCAGTATACCAAGCTCAGACCACT 817  
721 AGTTACCCACCCCACTGATGTCAGCAGCCCACTGATGTCAGTATACCAAGCTCAGACCACT 780  
818 AGCAGCTACGCGCAGCAGTATGAGCAGCCCACTGATGTCAGTATACCAAGCTCAGACCACT 877  
781 AGCAGCTACGCGCAGCAGTATGAGCAGCCCACTGATGTCAGTATACCAAGCTCAGACCACT 840

QY 878 GGGCAGAGAGCTTGGAGAGATTTTCCGACCCAGAGAGAAACCGGAGCATGAGTGGCCCTGAT 937  
 DB 841 GGGCAGAGAGCTTGGAGAGATTTTCCGACCCAGAGAGAAACCGGAGCATGAGTGGCCCTGAT 900  
 QY 938 AACCCGGGAGAGGGGAGAGGAGATTTGATCTGTGAGAGCATGAGAGAGTGGCGGGAG 997  
 DB 901 AACCCGGGAGAGGGGAGAGGAGATTTGATCTGTGAGAGCATGAGAGAGTGGCGGGAG 960  
 QY 998 GAGAGACGCGGTGAAATGGGACGCGTGGAGACGAGTGGCTTCAATTAAGCTGTGTGA 1057  
 DB 961 GAGAGACGCGGTGAAATGGGACGCGTGGAGACGAGTGGCTTCAATTAAGCTGTGTGA 1017  
 QY 1058 CCCATGAGATGAGACCAAGATCTTGAATCTAGGCGCCACTGATGATCAAGTGAAGCTCT 1117  
 DB 1018 CCCATGAGATGAGACCAAGATCTTGAATCTAGGCGCCACTGATGATCAAGTGAAGCTCT 1077  
 QY 1118 GACAAAGATGCAATTTATGTACAAAGATTTAATGACAGTGTGACTGATGATCTGGA 1177  
 DB 1078 GACAAAGATGCAATTTATGTACAAAGATTTAATGACAGTGTGACTGATGATCTGGA 1137  
 QY 1178 GACTTCTTTAAGCAGTGTGGGTTGTTAAGATGACAGAACTGGGCAACCATGATC 1237  
 DB 1138 GACTTCTTTAAGCAGTGTGGGTTGTTAAGATGACAGAACTGGGCAACCATGATC 1197  
 QY 1238 CAAATCTAAGTGTGACAAAGAAACAGAGAAACCCAAAGGCGATGCCAAGTGTCTATGA 1297  
 DB 1198 CAAATCTAAGTGTGACAAAGAAACAGAGAAACCCAAAGGCGATGCCAAGTGTCTATGA 1257  
 QY 1298 GACCCACCCACTGCAAGGCTGCGCTGAGATGGTTGATGGGAAAGATTTTCAAGGAGAC 1357  
 DB 1258 GACCCACCCACTGCAAGGCTGCGCTGAGATGGTTGATGGGAAAGATTTTCAAGGAGAC 1317  
 QY 1358 AAACTTAAGTCTTCTCTGCTCGAGAGAGCTTCAATGAACGATGCGGGGTGCTG 1417  
 DB 1318 AAACTTAAGTCTTCTCTGCTCGAGAGAGCTTCAATGAACGATGCGGGGTGCTG 1377  
 QY 1418 CCAACCCCGTGAAGGAG 1477  
 DB 1378 CCAACCCCGTGAAGGAG 1437  
 QY 1478 GGTCTGAGGAG 1537  
 DB 1438 GGTCTGAGGAG 1497  
 QY 1538 CCAAG 1597  
 DB 1498 CCAAG 1557  
 QY 1598 GGTGAG 1657  
 DB 1558 GGTGAG 1617  
 QY 1658 GAGTGAACCAAGTGAAG 1717  
 DB 1618 GAGTGAACCAAGTGAAG 1677  
 QY 1718 CCGGAGTGTGATGCTGAGCAAGAGTGGCTGAGTGGAGAGAGAGAGAGAGAGAGAGAGAG 1777  
 DB 1678 CCGGAGTGTGATGCTGAGCAAGAGTGGCTGAGTGGAGAGAGAGAGAGAGAGAGAGAGAG 1737  
 QY 1778 ATGAT 1837  
 DB 1738 ATGAT 1797  
 QY 1838 TTCCTGAT 1897  
 DB 1798 TTCCTGAT 1857  
 QY 1898 GGG 1957  
 DB 1858 GGG 1917

QY 1958 CCTGAAAAAATGATTAAGGCGAGAGACCGTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2017  
 DB 1918 CCTGAAAAAATGATTAAGGCGAGAGACCGTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1977  
 QY 2018 CAGAGACCCCGAGAGCTGAGATGACATCCAGATTTATTTTAAACGAGAAATGTTTT 2077  
 DB 1978 CAGAGACCCCGAGAGCTGAGATGACATCCAGATTTATTTTAAACGAGAAATGTTTT 2037  
 QY 2078 AAATTTAATTTCCATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTT 2137  
 DB 2038 AAATTTAATTTCCATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTT 2097  
 QY 2138 CTTTATGATTTTTCACCATTTGTGAGAAACATTTAAACAGTTAAATGTA 2189  
 DB 2098 CTTTATGATTTTTCACCATTTGTGAGAAACATTTAAACAGTTAAATGTA 2149  
 RESULT 11  
 AK056309  
 LOCUS  
 DEFINITION  
 Homo sapiens cDNA FLJ11747 f19, clone NR21007377, highly similar to RNA-BINDING PROTEIN EWS.  
 ACCESSION  
 AK056309  
 VERSION  
 AK056309.1 GI:16551673  
 KEYWORDS  
 oligo cloning; fls (full insert sequence).  
 SOURCE  
 Homo sapiens (human)  
 ORGANISM  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
 REFERENCE  
 AUTHORS  
 Oca, T., Suzuki, Y., Nishikawa, T., Otsuki, T., Sugiyama, T., Irie, R., Wakamatsu, A., Hayashi, K., Sato, H., Nagai, K., Kimura, K., Makita, H., Sekine, M., Obayashi, M., Nishi, T., Shibahara, T., Tanaka, T., Ishii, S., Yamamoto, J., Saito, K., Kawai, Y., Isono, Y., Nakamura, Y., Nagahara, K., Murakami, K., Yasuda, T., Iwayanagi, T., Wagatsuma, M., Shiratori, A., Sudo, H., Hosoliri, T., Kaku, Y., Kodaira, H., Kondo, H., Sugawara, M., Takahashi, M., Kanda, K., Yokoi, T., Furuya, T., Kikkawa, E., Omura, Y., Abe, K., Kamihara, K., Katsuta, N., Sato, K., Tanikawa, M., Yamazaki, M., Niinomiya, K., Ishihashi, T., Yamashita, H., Murakawa, K., Fujimori, K., Tanai, H., Kimura, M., Watanabe, M., Hirooka, S., Chiba, Y., Ishida, S., Ono, Y., Takiguchi, S., Watanabe, S., Yosida, M., Houchi, T., Kusano, J., Kanehori, K., Takahashi, Fujii, A., Hara, H., Tanase, T., Nomura, Y., Togiya, S., Komai, F., Hara, R., Takeuchi, K., Arita, M., Imose, N., Mitsuhashi, K., Yuki, H., Oshima, A., Saeki, N., Aotsuka, S., Yoshikawa, Y., Matsunawa, H., Ichihara, T., Shiohata, N., Sano, S., Moriya, S., Momiyama, H., Satoh, N., Takami, S., Terashima, Y., Suzuki, O., Nakagawa, S., Senoh, A., Mizoguchi, H., Goto, Y., Shimizu, F., Wakebe, H., Hishigaki, H., Watanabe, T., Sugiyama, A., Takemoto, M., Kawakami, B., Yamazaki, M., Watanabe, K., Kumsagi, A., Itakura, S., Fukuzumi, Y., Fujimori, Y., Komiyama, M., Tashiro, H., Tanigami, A., Fujiwara, T., Ono, T., Yamada, K., Fujii, Y., Ozaki, K., Hiro, M., Ohmori, Y., Kawabata, A., Hiki, T., Kobatake, N., Inagaki, H., Ikema, Y., Okamoto, S., Ohtani, R., Kawakami, T., Noguchi, S., Itoh, T., Shigeta, K., Senba, T., Matsumura, K., Nakajima, Y., Mizuno, T., Morinaga, M., Saeki, M., Togashi, T., Oyama, M., Hara, H., Watanabe, M., Komatsu, T., Mizushima-Sugano, J., Satoh, T., Shirai, Y., Takahashi, Y., Nakagawa, K., Okumura, K., Nagase, T., Nomura, N., Kikuchi, H., Masuno, Y., Yamashita, R., Nakai, K., Yada, T., Nakamura, Y., Ohara, O., Isegaki, T., and Sugano, S. Complete sequencing and characterization of 21,243 full-length human cDNAs  
 Nat. Genet. 36 (1), 40-45 (2004)  
 JOURNAL  
 PUBLISHED  
 REFERENCE  
 AUTHORS



QY 1445 CCAACACTCCGTGGAGGTCCAGAGAGCCAGAGAGTCTCTGGGGAGACCATGGATCCCATG 1504  
 DB 1438 CCACACTCCGTGGAGGTCCAGAGAGCCAGAGAGTCTCTGGGGAGACCATGGATCCCATG 1497  
 QY 1505 GAGAGCGGTGGAGAGATAGAGAGGCTTCCCTCAAGAGAGCCCGGGGTTCCAGAGG 1564  
 DB 1498 GAGAGCGGTGGAGAGATAGAGAGGCTTCCCTCAAGAGAGCCCGGGGTTCCAGAGG 1557  
 QY 1565 AACCCCTCTGAGAGAGAAAGTCCAGACCCAGAGCTGGAGACTGGGAGTCTCCCATCCG 1624  
 DB 1558 AACCCCTCTGAGAGAGAAAGTCCAGACCCAGAGCTGGAGACTGGGAGTCTCCCATCCG 1617  
 QY 1625 GGTGTGGAAACCAAGACTTCCCTGGAGAAACAGAGTCAACAGGTATAGGCCCAAG 1684  
 DB 1618 GGTGTGGAAACCAAGACTTCCCTGGAGAAACAGAGTCAACAGGTATAGGCCCAAG 1677  
 QY 1685 CTTGAAGGCTTCTCCCGCAACCTTCCGCGCGGGTGGTANTCTGGCAGAGGTGGC 1744  
 DB 1678 CTTGAAGGCTTCTCCCGCAACCTTCCGCGCGGGTGGTANTCTGGCAGAGGTGGC 1737  
 QY 1745 CTTGTGGCAATGGCGGAGAGAAAGTGGCTTATGATGATGATGATGATGATGATGATG 1804  
 DB 1738 CTTGTGGCAATGGCGGAGAGAAAGTGGCTTATGATGATGATGATGATGATGATGATG 1797  
 QY 1805 TTGAGAGGTGGCGGTGGTGGAGACAGAGGTGGCTTCCGTGGCGCGGGCATGGACCGA 1864  
 DB 1798 TTGAGAGGTGGCGGTGGTGGAGACAGAGGTGGCTTCCGTGGCGCGGGCATGGACCGA 1857  
 QY 1865 GGTGGCTTGTGGAGAGAGACAGAGGTGGCTTGGGGGGGCGCCCTGGACCTTTGATGAA 1924  
 DB 1858 GGTGGCTTGTGGAGAGAGACAGAGGTGGCTTGGGGGGGCGCCCTGGACCTTTGATGAA 1917  
 QY 1925 CAGATGGAG 1984  
 DB 1918 CAGATGGAG 1977  
 QY 1985 GGTGAG 2044  
 DB 1978 GGTGAG 2037  
 QY 2045 ACCAGATTTATTTTAAACAGAGAAATGTTTAAATTTAAATTTAAATTTAAATTTAAAT 2104  
 DB 2038 ACCAGATTTATTTTAAACAGAGAAATGTTTAAATTTAAATTTAAATTTAAATTTAAAT 2097  
 QY 2105 TGGCACAACAATTATGATTTCTTGTCTGACTTATGATTTTTCACATTGTGAAG 2164  
 DB 2098 TGGCACAACAATTATGATTTCTTGTCTGACTTATGATTTTTCACATTGTGAAG 2157  
 QY 2165 AAACATTTAAACAAGTTAAATGTA 2189  
 DB 2158 AAACATTTAAACAAGTTAAATGTA 2182  
 RESULT 12  
 AC137500/c 193559 bp DNA linear HTG 20-NOV-2002  
 LOCUS Homo sapiens chromosome 16 clone RP11-343G17, WORKING DRAFT  
 DEFINITION  
 AC137500  
 ACCESSION  
 AC137500.1 GI:25139894  
 VERSION  
 HTG: HTGS PHASE1; HTGS\_DRAFT; HTGS\_ACTIVEFIN.  
 KEYWORDS  
 Homo sapiens (human)  
 ORGANISM  
 Homo sapiens  
 Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 Unpublished  
 2 (bases 1 to 193559)  
 REFERENCE  
 DOB Joint Genome Institute.  
 AUTHORS  
 TITLE  
 Direct Submission

JOURNAL  
 Submitted (20-NOV-2002) Production Sequencing Facility, DOE Joint  
 Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
 COMMENT  
 -----Genome Center  
 Center: Joint Genome Institute  
 Center Code: JGI  
 Web site: <http://www.jgi.doe.gov>  
 -----  
 Project Information  
 Center Project Name: 538238  
 Center clone name: RP11-343G17  
 -----  
 Summary Statistics  
 Consensus quality: 192718 bases at least Q40  
 Consensus quality: 192915 bases at least Q30  
 Consensus quality: 193071 bases at least Q20  
 Estimated insert size: 160000; agarose-fp estimation  
 Estimated insert size: 193459; sum-of-contigs estimation  
 Quality coverage: 12.34 in Q20 bases; agarose-fp estimation  
 Quality coverage: 10.21 in Q20 bases; sum-of-contigs estimation.  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 2 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.  
 \* 1 96955: contig of 96955 bp in length  
 \* 96956 97055: gap of unknown length  
 \* 97056 193559: contig of 96504 bp in length.  
 FEATURES  
 source  
 1..193559  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"  
 /chromosome="16"  
 /clone="RP11-343G17"  
 /clone\_lib="RP11 human BAC library 11"  
 ORIGIN  
 Query Match 85.3%; Score 2038.4; DB 2; Length 193559;  
 Best Local Similarity 96.4%; Pred. No. 0;  
 Matches 2106; Conservative 0; Mismatches 76; Indels 2; Gaps 2;  
 QY 6 GAGAGGAGGTTGAG 65  
 DB 31044 GATGGGAGAGTTGAG 30985  
 QY 66 CCTATAGCAAGCTGAG 125  
 DB 30984 CCTATAGCAAGCTGAG 30925  
 QY 126 GATATGACAGACCAACCAAGGATATGGCAACAAAGCTATGGAACCTATGACAGCCCA 185  
 DB 30924 GATATGACAGACCAACCAAGGATATGGCAACAAAGCTATGGAACCTATGACAGCCCA 30865  
 QY 186 CTGATGTGAGCTTATACCAAGGCTGAGACCAAGCTGAGACCAAGCTGAGACCAAGCTGAG 245  
 DB 30864 TTGATGTGAGCTTATACCAAGGCTGAGACCAAGCTGAGACCAAGCTGAGACCAAGCTGAG 30805  
 QY 246 CTTCTATGAGAGAGCTTCCCACTGTTATATCACTCACTGAGCCCGCCAGGAGATAGAGCC 305  
 DB 30804 CTTCTATGAGAGAGCTTCCCACTGTTATATCACTCACTGAGCCCGCCAGGAGATAGAGCC 30745  
 QY 306 AGCTGTCCAGAGGATATGAG 365  
 DB 30744 AGCTGTCCAGAGGATATGAG 30685  
 QY 366 CCAAGGCTTCTATGAG 425  
 DB 30684 CCAAGGCTTCTATGAG 30625  
 QY 426 GCGAGAGCGAGAGCACTGACCTACAAAGACCGAGATGGAACCAAGCCACTGAGA 485





Db	10349	GATGGGGAACGTTGAGAAACGAGAGAAAGAGAAAATGCGATCTCATGATTCACGTA	10408
Qy	66	CCATATAGCCAGCTGCGAGGGCGACGGGCTTAAGTCTTAAACGCCGCCACCTCAAG	125
Db	10409	CCATATAGCCAGCTGCGAGGGCGACGGGCTTAAGTCTTAAACGCCGCCACCTCAAG	10468
Qy	126	GATATGCAAGAACCCAGGCGATATGCGCAACAAAGCTATGGAACCTATGAGACGCCA	185
Db	10469	GATATGCAAGAACCCAGGCGATATGCGCAACAAAGCTATGGAACCTATGAGACGCCA	10528
Qy	186	CTGATGTCAGCTATTAACCAAGGCTCAGACCACTGCAACCTATGGGCGAGACCGCTATGCA	245
Db	10529	TTGATGTCAGCTATTAACCAAGGCTCAGACCACTGCAATCTATGGGCGAGACCGCTATGCA	10588
Qy	246	CTTCTTATGAGACGCTCCCACTGGTATATCTATCTTCAACCTGCCCCCAGGAGTACAGCC	305
Db	10589	CTTCTTATGAGACGCTCCCACTGGTATATCTATCTTCAACCTGCCCCCAGGAGTACAGCC	10648
Qy	306	AGCCTGCGAGGGGATATGAGCACTGGGCTTATGATATCCACACTGCTACTAGTCAACCA	365
Db	10649	AGCCTGTCAGGGGATATGAGCACTGGGCTTATGATATCCACACTGCTACTAGTCAACCA	10708
Qy	366	CCGAGGCTCTTATGAGAGCTCAGTCTGCAATATGCACTGAGCTGCTTATCCAGGCTATG	425
Db	10709	CCGAGGCTCTTATGAGAGCTCAGTCTGCAATATGCACTGAGCTGCTTATCCAGGCTATG	10768
Qy	426	GGCAGCAGCCAGCAGCACTGCACTCTTCAAGACCGGAGATGGAACCAAGCCCACTGAGA	485
Db	10769	GGCAGCAGCCAGTATGAGCACTGCACTCTTCAAGACTGAGATGGAACCAAGCCCACTGAGA	10828
Qy	486	CTAATCAACCTCAATCTATAGACAGGGGGGTTACAAACAGGCCAGCCTAGGATATGAGACAGA	545
Db	10829	CTAATCAACCTCAATCTATAGACAGGGGGGTTACAAACAGGCCAGCCTAGGATATGAGACAGA	10888
Qy	546	GTAATCTACAGTTATCCCAAGGTACCTGGAGGCTACCCCATGAGCCAGTCACTGACCTC	605
Db	10889	GTAATCTAGATTATCCCAAGGTATCTGGAGGCTACCATGAGCCAGTCACTGACCTC	10948
Qy	606	CATCTTACCTCTTACCAAGCTATTTCTTCTTACACAGCCCACTATGATATCAGACAGTT	665
Db	10949	CATCTTACCTCTTACCAAGCTATTTCTTCTTACACAGCCCACTATGATATCAGACAGTT	11008
Qy	666	ACTCTGAGCAAAACCTTATGGGCAACCGAGAGCTATGAGACAGCAGATAGCTATGCTC	725
Db	11009	ACTCTGAGCAAAACCTTATGGGCAAAACCGAGAGCTATGAGACAGCAGATAGCTATGCTC	11068
Qy	726	AACAAACAGCTATGCGGACAGACCTTCCCACTAGTTACCAACCCCAACTGATCTTACA	785
Db	11069	AACAAACAGCTATGCGGACAGCTTCCCACTAGTTACCAACCCCAACTGATCTTACA	11128
Qy	786	GCCAAAGCTCCAGTCAATATATGCGAACAGACGACGACTATCGGGCAGCAGATTCATTC	845
Db	11129	GCCAAAGCTCCAGTCAATATATGCGAACAGACGACGACTATCGGGCAGCAGATTCATTC	11188
Qy	846	GACAGGACCAACCCAGTAGATGGGGTTATATGGGAGAGATCTGGAGATTTTCCGAGC	905
Db	11189	GACAGGACCAACCCAGTAGATGGGGTTATATGGGAGAGATCTGGAGATTTTCCGAGC	11248
Qy	906	CAGGAGGAACCCGAGCATGATGTCCTCTGATTAACCGGGGCAAGGGGAGAGGGGAGATTGG	965
Db	11249	TAGGAGGAACCCGAGCATGATGTCCTCTGATTAACCGGGGCAAGGGGAGAGGGGAGATTGG	11308
Qy	966	ATCTGTGAGGCATATGACAGAGGTGGGCGGGAGAGAGACGCGGTGGAATGGGCAAGGCTTG	1025
Db	11309	ATCTGTGAGGCATATGACAGAGGTGGGCGGGAGAGAGATGCGGTGGAATGGGCAAGGCTTG	11368
Qy	1026	GAGAGCGAGGTGGCTTCAATTAACCTGTGTGAGACCCATGGAATGAAGACCAAGATCTTGATC	1085
Db	11369	GAGAGCGAGGTGGCTTCAATTAACCTGTGTGAGACCCATGGAATGAAGACCAAGATCTTGATC	11428
Qy	1086	TAGGCGCAACCTGATATCCAGATGAAGAATCTGACAAACAGTGCAATTTATGTACAGAGAT	1145
Db	11429	TAGGCGCAACCTGATATCCAGATGAAGAATCTGACAAACAGTGCAATTTATGTACAGAGAT	11488



QY	1146	TAAATGACAGTGTGCTCTAGATGATCTGGCAGACCTCTTTAAGCAGTGTGGGTTGTGA	1205
Db	11489	TAAATGACAACTGTGACTCTAGATGATCTGGTAGACTCTTTAAGCAGTGTGGGTTGTGA	11548
QY	1206	AGATGAACAAGAGAATTGGGCAACCCATGATCCACATCTACCTGACAAAGAAACAGAA	1265
Db	11549	AGATGAACAAGAGAATTGAGCAACCCATGATCCACATCTACCTGACAAAGAAACAAGAA	11608
QY	1266	AGCCCAAAAGGGATGCCCAACATGTCTCTATGAAGCCCACTGCCCAGGCTGGCCGTGG	1325
Db	11609	AGCCCAAAAGTGAATGCCCAACATGTCTCTGTGAAGCTCACTACCTGCCAAAGCTGCCGTGG	11668
QY	1336	AATGTTTGTATGGAAAGATTTTCAAGGAGCAAACTTAAAGTCTCCCTTGCTCGGAGA	1385
Db	11669	AATGTTTGTATGGAAAGATTTTCAAGGAGCAAACTTAAAGTCTCTTGCTCGGAGA	11728
QY	1386	AGCTCTCCAAATGAACGTATGCGGGGTGTCTGCCACCCCGTAGGGGAGAGCATGCCAC	1445
Db	11729	GGCTCTCCAGTAAACGATATGCAAGGATGGTATGCAACCCCATGAGGGCAGAGGATCCAC	11788
QY	1446	CACCACTCCGTGAGAGTCTCAGAGAGCCCAAGAGGTCTGGGGAGACCATAGGTGGATGG	1505
Db	11789	CACCACTCTGCGGAGGTCTCAGAGAGCCCAAGAGGTCTGGGGAGACCATAGGTGCATGG	11848
QY	1506	GAGGCCGTGAGAGATATAGAGAGCTTCCCTCCAAAGAGACC CGGGTTTCCGAGGGA	1565
Db	11849	GAGGCCGTGAGAGATATAGAGAGCTTCCCTCCAAAGAGACC CGGGTTTCCGAGGGA	11908
QY	1566	ACCCCTCTGAGAGAGGAAACGTCCAGACCCGAGTGTGAGACTGGCATGTCCCATCCGG	1625
Db	11909	ACCCCTCTGAGAGAGGAAACGTCCAGACCCAGCTGTGAGACAGCATGTCTCCAAATCCGG	11968
QY	1626	GTTGTGTGAACACAGAACTTGCGCTGTGAGAAACAGAGTGCACACAGTGTAAAGCCCCAAAGC	1685
Db	11969	GTTGTGTGAACACAGAACTTGCGCTGTGAGAAACAGAGAGCAACAGTGTAAAGCTCCAAAGC	12028
QY	1686	CTGAAGGCTTCTCTCCCGCACCTTTTCGCCCCCGGGTGTATCTGTGCAGAGTGGCC	1745
Db	12029	CTGAAGGCTTCTCTCCCGCACCTTTTCGCCCCGGGTGTATCTATGACAGAGTGGCC	12088
QY	1746	CTGTGTGCATTCGSGGAGGAAAGAGTGGGCTCATATGATGATGCTCCCGTGAATGT	1805
Db	12089	CTGTGTGCATATGSGGAGGAAAGAGTGGGCTCATATGATCATATGATGCTCCCGTGAATGT	12148
QY	1806	TCAGAGGTGGCCGTGTGTGAGAACAGAGGTGGCTTCCGTGTGGCCGGGCAATGCACCGAG	1865
Db	12149	TCAGAGGTGGCTGTGTGTGAGAACAGAGGTGGCTTCCGTGTGGCTGGGCAATGCACCGAG	12207
QY	1866	GTTGCTTTTGTGTGAGAAAGACAGAGTGGCCCTTGGGGGGCCCTCTGACCTTTGATGAAAC	1925
Db	12208	GTTGCTTTTGTGTGAGAAAGACAGAGTGGCCCTTGGGGGGCCCCC-GGACCTTTGATGATTAAC	12266
QY	1926	AGATGGGAGGAAAGAGAGAGACGTGTGAGACCTGTGAAAAATGTATTAAGGCGAGCACCC	1985
Db	12267	CAATGGGAGGAAAGAGAGAGACGTGTGAGACCTGTGAAAAACGATTAAGGCGAGCACT	12326
QY	1986	GTCAGAGCGCAGAGATGCGCCCTTACTAGATGCAGAGACCCCGCAGAGCTGATGACTTA	2045
Db	12327	GTCAGAGCGCAGAGATCAACCTTACTAGATGCAGAGAACCCCGCAGAGCTGATGACTTA	12386
QY	2046	CCAGATTTATTTTAAACAGAAAAATGTTTAAATTTAATTAATTCATTTATATATGT	2105
Db	12387	CCAGATTTATTTTAAACAGAAAAATGTTTAAATTTAATTAATTCATTTATATATGT	12446
QY	2106	GGCCACAACATATGATTAATCTGTGTCTGTACTTTAGATATTTTCCATTTGTGAAGA	2165
Db	12447	GGCCACAACATATGATTAATCTTTTCTCTGTACTTTAGATATTTTTCACCATTTGTGAAGA	12506
QY	2166	AACATTTAAACAAATTAAATGTGA	2189
Db	12507	AACATTTAAACAAATTAAATGTGA	12530

```

RESULT 14
AL596087/c 180718 bp DNA linear PRI 10-MAR-2002
LOCUS
DEFINITION Human DNA sequence from clone RP11-9L18 on chromosome 1, complete
sequence.
ACCESSION
VERSION AL596087
KEYWORDS
SOURCE AL596087..11 GI:19351911
ORGANISM
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniota; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 180718)
Almeida,J.
REFERENCE
AUTHORS Direct Submission
TITLE Submitted (10-MAR-2002) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humangear@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
On Mar 11, 2002 this sequence version replaced g5:16944185.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. The following
abbreviations are used to associate primary accession numbers given
in the feature table with their source databases: Em:, EMBL; Sw:,
SWISSPROT; Tr:, TrEMBL; Wp:, WormPeP; Information on the WormPeP
database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
was generated from part of bacterial clone contigs of human
chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping
Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr1
RP11-9L18 is from the library RPCI-11.1 constructed by the group of
Pleier de Jong. For further details see
http://www.chori.org/bacpac/home.htm
VECTOR: pBAC3.6
This sequence is the entire insert of clone RP11-9L18 The true left
end of clone RP11-47907 is at 136234 in this sequence. The true
right end of clone RP11-375H19 is at 73742 in this sequence.
location/Qualifiers
1. 180718
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="1"
/clone="RP11-9L18"
/clone_1fb="RPCI-11.1"
23201..23229
/note="Sequence confirmed by AC011221 and AC018774
sequenced by W18R."
misc_feature (23230..23297)
/note="[ ] Unsure [x] Misc feature Then select the text
for the note(s) - [ ] Tandem repeat [ ] Single clone
region [ ] Forced join [x] Other Add a comment here -
Sequence from AC011221 and AC018774 sequenced by W18R."
misc_feature Complement(46397..46442)
/note="Single clone region. Sequence from reads from a
short insert library derived from a single pUC clone.
Restriction digest data confirm the assembly."
misc_feature Complement(102116..102124)
/note="767 bases of ISI transposon (X52534) removed here.
This sequence represents the duplicated flanking sequence
of the ISI."

```



OY	2106	GGCCACAAATATGATGTATTCTCGTGCAGTAATTAGATTTTTCACCATTTGCAAGA	2165
DB	155352	GCCACACATTAATGATTAATCTTCTCTGACTTAATGATTTTTCACCATTGTGAAG	155473
OY	2166	AACATTAANAACAAGTTAATGTA	2189
DB	155472	AACATTTAAAACAAGTTAATGTA	155449
RESULT 15			
AC011221		182501 bp	DNA
LOCUS		11near	HTG 04-SEP-2000
DEFINITION		Homio sapiens clone RP11-9L18, WORKING DRAFT SEQUENCE, 13 unordered pieces.	
ACCESSION		AC011221	
VERSION		AC011221.5	GI:9966283
KEYWORDS		HTG; HTGS PHASE1; HTGS_DRAFT.	
SOURCE		Homio sapiens (human)	
ORGANISM		Homio sapiens	
REFERENCE		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiindae; Homo.	
AUTHORS		1 (bases 1 to 182501)	
TITLE		Birren,B., Linton,L., Nusbaum,C. and Lander,E.	
JOURNAL		Homio sapiens, clone RP11-9L18	
REFERENCE		Unpublished	
AUTHORS		2 (bases 1 to 182501)	
TITLE		Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M., Baldwin,J., Barne,N., Beckeily,R., Boguslavsky,L., Bouhgalter,B., Brown,A., Casale,A., Colangelo,M., Collins,S., Collymore,A., Cooke,P., Deatellano,K., Dewar,K., Domingo,M., Donelan,L., Doyle,M., Pereira,P., Fitzhugh,W., Forrest,C., Funke,R., Gage,D., Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heatord,A., Horton,L., Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,T., Lehoczy,T., Lien,C., Locke,K., McDonald,P., Marquis,N., McEwan,P., McGuirk,A., McKernan,C.H., McLaughlin,J., Meldrum,J., Morrow,J., Naylor,J., Norman,K., O'Connor,F., O'Donnell,P., Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P., Stange-Thomann,N., Stefanovic,N., Subramanian,A., Talamas,J., Testave,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,T., Wu,X., Wymen,D., Ye,W.J., Zimmer,A. and Zody,M.	
COMMENT		Direct Submission	
TITLE		Submitted (03-OCT-1999) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02131, USA	
JOURNAL		On Sep 4, 2000 this sequence version replaced GI:7341864.	
ALL repeats were identified using RepeatMasker:			
Smitt, A.P.A. & Green, P. (1996-1997)			
http://ftp.genome.washington.edu/RM/RepeatMasker.html			
----- Genome Center -----			
Center: Whitehead Institute/ MIT Center for Genome Research			
Center code: WIBR			
Web site: http://www-seq.wi.mit.edu			
Contact: sequence.submissions@genome.wi.mit.edu			
----- Project Information -----			
Center project name: L3016			
Center Clone name: 9_L_18			
----- Summary Statistics -----			
Sequencing vector: MJ3; M7781S; 100% of reads			
Chemistry: Dye-terminator Big Dye; 100% of reads			
Assembly program: Phrap; version 0.960731			
Consensus quality: 175883 bases at least Q40			
Consensus quality: 179186 bases at least Q30			
Consensus quality: 180361 bases at least Q20			
Insert size: 182000; agarose-fp			
Insert size: 181301; sum-of-contigs			
Quality coverage: 5.1 in Q20 bases; agarose-fp			
Quality coverage: 5.1 in Q20 bases; sum-of-contigs			
* NOTE: This is a 'working draft' sequence. It currently consists of 13 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as * runs of N, but the exact sizes of the gaps are unknown. * This record will be updated with the finished sequence.			

```

* as soon as it is available and the accession number will
* be preserved.
1
6942: contig of 6941 bp in length
*
* 6942 7041: gap of 100 bp
*
* 7042 8102: contig of 1061 bp in length
*
* 8103 8202: gap of 100 bp
*
* 8203 9606: contig of 1404 bp in length
*
* 9607 9707: gap of 100 bp
*
* 9707 13243: contig of 3537 bp in length
*
* 13244 13343: gap of 100 bp
*
* 13344 19155: contig of 5812 bp in length
*
* 19156 19255: gap of 100 bp
*
* 19256 28050: contig of 8795 bp in length
*
* 28051 28150: gap of 100 bp
*
* 28151 39485: contig of 11334 bp in length
*
* 39485 39585: gap of 100 bp
*
* 39585 54855: contig of 15271 bp in length
*
* 54856 54956: gap of 100 bp
*
* 54956 68228: contig of 13273 bp in length
*
* 68229 68328: gap of 100 bp
*
* 68329 88542: contig of 20214 bp in length
*
* 88543 88642: gap of 100 bp
*
* 88643 109342: contig of 20700 bp in length
*
* 109343 136118: gap of 100 bp
*
* 109943 136118: contig of 26676 bp in length
*
* 136119 136219: gap of 100 bp
*
* 136219 182501: contig of 46283 bp in length.
FEATURES
source
1..182501
location/Qualifiers
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/clone_id="RP11-9L118"
/clone_lib="RPCI-11 Human Male BAC"
1..6941
/note="assembly_fragment"
clone_end:SP6
vector_side:left"
7042..8102
/note="assembly_fragment"
8203..9606
/note="assembly_fragment"
9707..13243
/note="assembly_fragment"
13344..19155
/note="assembly_fragment"
19256..28050
/note="assembly_fragment"
28151..39484
/note="assembly_fragment"
39585..54855
/note="assembly_fragment"
54956..68228
/note="assembly_fragment"
68329..88542
/note="assembly_fragment"
88643..109342
/note="assembly_fragment"
109443..136118
/note="assembly_fragment"
136219..182501
/note="assembly_fragment"
clone_end:T7
vector_side:right"
ORIGIN
Query Match 85.1%; Score 2033.6; DB 2; Length 182501;
Best Local Similarity 96.3%; Pred. No. 0;
Matches 2103; Conservative 0; Mismatches 79; Indels 2; Gaps 2;
6 GAGACGCACTTTAGAGCAACGAGAGAGAAAGAGAGAAAATGCGCTTCACGGATTACAGTA 65
|||
56444 GATGCGGACCTTATAGAGAAACGAGAGAGAAAGAGAGAAAATGCGCTTCATGATTACAGTA 56503

```

QY 66 CCTATAGCCAAAGCTGAGGCGAGGCGGTACAGTGCCTTACACCCGCCAGCCCACTCAAG 125  
| | | | |  
Db 56504 CCTATAGCCAAAGCTGAGGCGAGGCGGTACAGTGCCTTACACCCGCCAGCCCACTCAAG 56563  
QY 126 GATATGACAGACCAAGCCAGGCAATAGGCGCAACAAAGCTATGGAACCTATGACAGCCCA 185  
| | | | |  
Db 56564 GATATGACAGACCAAGCCAGGCAATAGGCGCAACAAAGCTATGGAACCTATGACAGCCCA 56623  
QY 186 CTGATGTCAGTATACCCAGGCTCAGACCACTGCAACCTATGGGCGAGACCCGCTTATGCA 245  
| | | | |  
Db 56624 TTGATGTCAGTATACCCAGGCTCAGACCACTGCAACCTATGGGCGAGACCCGCTTATGCA 56683  
QY 246 CTTCCTTATGAGACAGCTCCCACTGCTTATACCTACCTGCACTGCCCCCGGAGATCAGCC 305  
| | | | |  
Db 56684 CTTCCTTATGAGACAGCTCCCACTGCTTATACCTACCTGCACTGCCCCCGGAGATCAGCC 56743  
QY 306 AACCTGTCCAGGGGTATGCGACTGTGCTTATGATACCACTGCTACAGTCAACCA 365  
| | | | |  
Db 56744 AACCTGTCCAGGGGTATGCGACTGTGCTTATGATACCACTGCTACAGTCAACCA 56803  
QY 366 CCCAGGCTCTTATGCAAGCTCAGTCTGCAATGCGACTGACCTGCTTATCCAGCTTATG 425  
| | | | |  
Db 56804 CCCAGGCTCTTATGCAAGCTCAGTCTGCAATGCGACTGACCTGCTTATCCAGCTTATG 56863  
QY 426 GCGAGGAGCGAGGCGCACTGCACTTACAAAGCCGAGGTGGAACCAAGCCCACTGAGA 485  
| | | | |  
Db 56864 GCGAGGAGCGAGGCGCACTGCACTTACAAAGCTGCAAGATGGAACCAAGCCCACTGAGA 56923  
QY 486 CTATGTCACCTCAATCTACACAGGGGGTTTACACAGCCAGCCCTAGATATGACAGA 545  
| | | | |  
Db 56924 CTATGTCACCTCAATCTACACAGGGGGTTTACACAGCCAGCCCTAGATATGACAGA 56983  
QY 546 GTTACTACAGTTATCCCGAGGTACCTGGGAGCTTACCCCATGCGACCACTGCACTGCACTTC 605  
| | | | |  
Db 56984 GTTACTACAGTTATCCCGAGGTACCTGGGAGCTTACCCCATGCGACCACTGCACTGCACTTC 57043  
QY 606 CATCCCACTCCCTACAGCTATTTCTCTACACAGGCGGACTGATTTGATCAGAGAGTT 665  
| | | | |  
Db 57044 CATCCCACTCCCTACAGCTATTTCTCTACACAGGCGGACTGATTTGATCAGAGAGTT 57103  
QY 666 ACTCTGAGAGAACACTTATGAGGCAACGAGCAGCTATGACAGCAGAGTATGATGATGTC 725  
| | | | |  
Db 57104 ACTCTGAGAGAACACTTATGAGGCAACGAGCAGCTATGACAGCAGAGTATGATGATGTC 57163  
QY 726 AACAAAGCAGCTATGAGGCGAGCCTTCCACTTGTATACCACTCCCAACTGCACTGATCTTACA 785  
| | | | |  
Db 57164 AACAAAGCAGCTATGAGGCGAGCCTTCCACTTGTATACCACTCCCAACTGCACTGATCTTACA 57223  
QY 786 GCCAAGCTCCAACTCAATATGACCAAGAGCAGCAGCTACGAGGCGAGCAGAGTTCATTCC 845  
| | | | |  
Db 57224 GCCAAGCTCCAACTCAATATGACCAAGAGCAGCAGCTACGAGGCGAGCAGAGTTCATTCC 57283  
QY 846 GACAGAGCAACCCAGTATGAGTGGTGTATGAGGCGAGAGTCTGAGAGATTTTCCGAGAC 905  
| | | | |  
Db 57284 GACAGAGCAACCCAGTATGAGTGGTGTATGAGGCGAGAGTCTGAGAGATTTTCCGAGAC 57343  
QY 906 CAGGAGAGAACCCGAGCAGTATGAGTGGTGTATGAGGCGAGAGTCTGAGAGATTTTCCGAGAC 965  
| | | | |  
Db 57344 CAGGAGAGAACCCGAGCAGTATGAGTGGTGTATGAGGCGAGAGTCTGAGAGATTTTCCGAGAC 57403  
QY 966 ATGCTGAGAGCAGTATGAGTGGTGTATGAGGCGAGAGTCTGAGAGATTTTCCGAGAGCTG 1025  
| | | | |  
Db 57404 ATGCTGAGAGCAGTATGAGTGGTGTATGAGGCGAGAGTCTGAGAGATTTTCCGAGAGCTG 57463  
QY 1026 GAGAGCAGAGTGGCTTCAATATGAGTGGTGTATGAGGCGAGAGTCTGAGAGATTTTCCGAGAG 1085  
| | | | |  
Db 57464 GAGAGCAGAGTGGCTTCAATATGAGTGGTGTATGAGGCGAGAGTCTGAGAGATTTTCCGAGAG 57523  
QY 1086 TAGGCCCACTGTATGATCCAGATGAGAGCTTGTACCAACAGTCAATTTATGTACAAAGAT 1145  
| | | | |  
Db 57524 TAGGCCCACTGTATGATCCAGATGAGAGCTTGTACCAACAGTCAATTTATGTACAAAGAT 57583

QY 1146 TAAATGACAGTGTACTATGATGATCTGAGAGCTTCTTTAAGCAGTGTGGGTTGTTA 1205  
| | | | |  
Db 57584 TAAATGACAGTGTACTATGATGATCTGAGAGCTTCTTTAAGCAGTGTGGGTTGTTA 57643  
QY 1206 AGATGACAGAGAACTGGGCAACCCATGATTCACATCTTACTGACAGAGAAACAGAA 1265  
| | | | |  
Db 57644 AGATGACAGAGAACTGGGCAACCCATGATTCACATCTTACTGACAGAGAAACAGAA 57703  
QY 1266 AGCCCAAGGGGATGCGCAGTGTCTATGAGAACCCCACTGCGCAAGGCTGCGTGG 1325  
| | | | |  
Db 57704 AGCCCAAGGGGATGCGCAGTGTCTATGAGAACCCCACTGCGCAAGGCTGCGTGG 57763  
QY 1326 AATGTTTATGAGGAAATTTTCAAGGAGCAACTTAAAGTCTCTCTGCTCGGAGAGA 1385  
| | | | |  
Db 57764 AATGTTTATGAGGAAATTTTCAAGGAGCAACTTAAAGTCTCTCTGCTCGGAGAGA 57823  
QY 1386 AACCTTCAATGACATATGCGGGGTGCTCTGCAACCCGTTGAGGCGAGAGGCAATGCCAC 1445  
| | | | |  
Db 57824 AACCTTCAATGACATATGCGGGGTGCTCTGCAACCCGTTGAGGCGAGAGGCAATGCCAC 57883  
QY 1446 CACCACTCCGTTGAGAGTCCAGGAGGCCAGAGAGTCTGAGGGGAGCCATGGGTGCGATGG 1505  
| | | | |  
Db 57884 CACCACTCCGTTGAGAGTCCAGGAGGCCAGAGAGTCTGAGGGGAGCCATGGGTGCGATGG 57943  
QY 1506 GAGGCGTGGAGAGATGAGAGAGGCTTCCCTCAAGAGAACCCGAGGTTCCGAGGGA 1565  
| | | | |  
Db 57944 GAGGCGTGGAGAGATGAGAGAGGCTTCCCTCAAGAGAACCCGAGGTTCCGAGGGA 58003  
QY 1566 ACCCTCTGAGAGAGAAAGTTCGACACCCGAGCTGAGAGCTGGCAGTGTCCCAATCCGG 1625  
| | | | |  
Db 58004 ACCCTCTGAGAGAGAAAGTTCGACACCCGAGCTGAGAGCTGGCAGTGTCCCAATCCGG 58063  
QY 1626 GTTGTGAAACAGAACTTCCGCTGAGAAACAGAGTGCACCAAGTGTAAAGCCCAAGC 1685  
| | | | |  
Db 58064 GTTGTGAAACAGAACTTCCGCTGAGAAACAGAGTGCACCAAGTGTAAAGCCCAAGC 58123  
QY 1686 CTGAAAGCTTCTCCCGCACCTTTCGCGCCCGGAGTGTATGCTGGCAGAGTGGCT 1745  
| | | | |  
Db 58124 CTGAAAGCTTCTCCCGCACCTTTCGCGCCCGGAGTGTATGCTGGCAGAGTGGCT 58183  
QY 1746 CTGTTGCGCATGCGGGAGAGAGAGTGGCTCATGATGTGTGTGTCTCCGCTGGAATGT 1805  
| | | | |  
Db 58184 CTGTTGCGCATGCGGGAGAGAGAGTGGCTCATGATGTGTGTGTCTCCGCTGGAATGT 58243  
QY 1806 TCAGAGGTGGCGGTGTGAGACAGAGGTGGCTTCGAGTGGCGCGGAGCTGAGACCGAG 1865  
| | | | |  
Db 58244 TCAGAGGTGGCGGTGTGAGACAGAGGTGGCTTCGAGTGGCGCGGAGCTGAGACCGAG 58302  
QY 58303 GTGGCTTGTGTGAGAGAAACAAAGGTGGCTTCGAGTGGCGCGGAGCTGAGACCGAG 58361  
| | | | |  
Db 58362 CATGTGAGAGAAAGAGAGAGACGTGAGAGACCTGGAAGAAACGGAATGAAGGAGACACT 58421  
QY 1986 GTTCAGAGGCGAGAGATGGGCTCTAGATGAGAGACCCGAGAGCTGAGTGAAGTGA 2045  
| | | | |  
Db 58422 GTTCAGAGGCGAGAGATGGGCTCTAGATGAGAGACCCGAGAGCTGAGTGAAGTGA 58481  
QY 2046 CCAGATTTATTTTAAACCAAGAAATGTTTAAATTTAATTTCAATTTAATTTAATTTAAT 2105  
| | | | |  
Db 58482 CCAGATTTATTTTAAACCAAGAAATGTTTAAATTTAATTTAATTTAATTTAATTTAATTTA 58541  
QY 2106 GGCACACATTTATGATTTATTCCTGTCTGATCTTTAGATTTTCAACATTTTGTGAAGA 2165  
| | | | |  
Db 58542 GGCACACATTTATGATTTATTCCTGTCTGATCTTTAGATTTTCAACATTTTGTGAAGA 58601  
QY 2166 AACATTTAAACAAATTAATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTAAT 2189  
| | | | |  
Db 58602 AACATTTAAACAAATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTAAT 58625

Tue Feb 22 11:31:10 2005

us-10-791-017a-1.rge

Page 25

Search completed: February 20, 2005, 11:50:21  
Job time : 10045.3 secs

3.1

***This Page Blank (uspto)***



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using bw model

Run on: February 20, 2005, 00:31:32 ; Search time 1159.11 Seconds  
(without alignments)  
12206.093 Million cell updates/sec

Title: US-10-791-017a-1

Perfect score: 2390  
1 gagggagagcgcgacttgcg.....taaaaaaaaaaaaaaaaa 2390

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq\_16Dec04:\*  
1: geneseqn1808:\*  
2: geneseqn1908:\*  
3: geneseqn2008:\*  
4: geneseqn2018:\*  
5: geneseqn2001b:\*  
6: geneseqn2002a:\*  
7: geneseqn2002b:\*  
8: geneseqn2003a:\*  
9: geneseqn2003b:\*  
10: geneseqn2003c:\*  
11: geneseqn2003d:\*  
12: geneseqn2004a:\*  
13: geneseqn2004b:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2390	100.0	2390	13	ADSI6277 Human CDN
2	2388.4	99.9	2390	6	ABK84628 Human CDN
3	2388.4	99.9	2390	6	ABK84628 Human CDN
4	2388.4	99.9	2390	6	ABK84628 Human CDN
5	2388.4	99.9	2390	13	ACN37537 Human Ewe
6	2369.4	99.1	2371	2	AAQ50643 Human Ewe
7	2357.8	98.7	2372	13	ADQ86032 Human Ewe
8	2281.2	95.4	2388	5	AA570647 Human Ewe
9	2179.6	91.2	2273	6	AA562623 Human Ewe
10	2141.8	89.6	2176	6	AA562623 Human Ewe
11	2033.4	85.1	2177	13	ACN37537 Human Ewe
12	1927.4	80.6	2188	6	AB193833 Human Ewe
13	1808.4	75.7	1988	10	ADQ53506 Human Ewe
14	1782.8	74.6	2026	13	ADQ53506 Human Ewe
15	1564.4	65.5	1807	13	ADQ53506 Human Ewe
16	1305	54.6	3309	2	AAQ50646 Human Ewe
17	819	34.3	2403	2	AAQ50646 Human Ewe
18	797.6	33.4	2412	2	AAQ50646 Human Ewe
19	758.4	31.7	856	13	ACN37540 Human Ewe

21	707.4	29.6	845	13	ACN37538	ACN37538 Tumour-as
22	653.6	27.3	1462	4	AA126725	AA126725 Human bre
23	570	23.8	2299	11	ACN88697	ACN88697 Breast ca
24	569.4	23.8	665	13	ADQ54544	ADQ54544 Novel can
25	514.6	21.5	540	4	AA117281	AA117281 Probe #72
26	514.6	21.5	540	4	ABA62106	ABA62106 Human foe
27	514.6	21.5	540	4	AA142050	AA142050 Probe #10
28	514.6	21.5	540	4	ABA29535	ABA29535 Probe #80
29	514.6	21.5	540	4	AAK36327	AAK36327 Human bon
30	514.6	21.5	540	4	AAK10423	AAK10423 Human bra
31	514.6	21.5	540	4	AB836007	AB836007 Human liv
32	514.6	21.5	540	6	AB810394	AB810394 Human gen
33	508.4	21.3	550	9	ACH40478	ACH40478 Human foe
34	488.4	20.4	684	13	ACN37539	ACN37539 Tumour-as
35	480.2	20.1	568	9	ACH22345	ACH22345 Human adu
36	435.4	18.2	468	9	ACH32497	ACH32497 Human end
37	431	18.0	454	2	AAV86051	AAV86051 EST clone
38	419.2	17.5	471	2	AA117153	AA117153 Probe #70
39	411.6	17.2	505	4	ABA61641	ABA61641 Human foe
40	411.6	17.2	505	4	AA141553	AA141553 Probe #10
41	411.6	17.2	505	4	ABA29300	ABA29300 Probe #77
42	411.6	17.2	505	4	AAK35836	AAK35836 Human bon
43	411.6	17.2	505	4	AAK09942	AAK09942 Human bra
44	411.6	17.2	505	4	AB835546	AB835546 Human liv
45	411.6	17.2	505	4	AB835546	AB835546 Human liv

## ALIGNMENTS

RESULT 1	ADSI6277	standard; cDNA; 2390 BP.
ID	ADSI6277	standard; cDNA; 2390 BP.
AC	ADSI6277;	
XX	XX	
DT	18-NOV-2004	(first entry)
DE	Human cDNA encoding the Ewing sarcoma protein SeqID 1.	
KW	human; ss; gene; Ewing sarcoma; EWS; prostatic cancer; alopecia; acne; hypogonadism; androgen-resistance syndrome; testicular feminisation.	
XX	XX	
OS	Homo sapiens.	
XX	XX	
FH	Key	Location/Qualifiers
FT	CDS	44..2014
FT		/tag= a
FT		/product= "Ewing sarcoma protein"
XX	XX	
XX	XX	
PN	EP1455190-A1.	
XX	XX	
PD	08-SEP-2004.	
XX	XX	
PF	16-FEB-2004; 2004EP-00003422.	
XX	XX	
PR	04-MAR-2003; 2003DE-01009280.	
XX	XX	
XX	25-APR-2003; 2003US-0465692P.	
PA	(SCHD ) SCHERING AG.	
XX	XX	
PI	Obendorf M, Wolf S;	
XX	XX	
DR	WPI; 2004-627861/61.	
DR	F-PSDB; ADSI6278.	
XX	XX	
PT	Determining the hormonal effects of substances, used to identify	
PT	pharmaceuticals, e.g. for treatment of androgen receptor dysfunction,	
PT	from modulating interaction between nuclear receptors and Ewing sarcoma	
XX	protein.	
PS	Claim 7; SEQ ID NO 1; 30pp; German.	
XX	XX	



Qy	1861	CCGAGGTGGCTTTGGTGGAGGAACAAGAGTGGCCCTGGGGGGGCCCTTCGACCTTTGAT	1920
Db	1861	CCGAGGTGGCTTTGGTGGAGGAACAAGAGTGGCCCTGGGGGGGCCCTTCGACCTTTGAT	1920
Qy	1921	GGAACAGATGGGAGGAGAAAGAGGAGGACGTGGAGGACCTGGAAAAATATGATTAAGGCGA	1980
Db	1921	GGAACAGATGGGAGGAGAAAGAGGAGGACGTGGAGGACCTGGAAAAATATGATTAAGGCGA	1980
Qy	1981	GCACCGTGAGAGCCGACAGATCGGCCCTACTAGATGACAGAGACCCGACAGCTGCATT	2040
Db	1981	GCACCGTGAGAGCCGACAGATCGGCCCTACTAGATGACAGAGACCCGACAGCTGCATT	2040
Qy	2041	GACTACCGAGTTTATTTTTTAAACCAAGAAAAATGTTTAAATTTATATTCATATTATA	2100
Db	2041	GACTACCGAGTTTATTTTTTAAACCAAGAAAAATGTTTAAATTTATATTCATATTATA	2100
Qy	2101	ATGTTGGCCACACATATATGATATATCTTGCTGTACTTTAGTATTTTCAACATTTGT	2160
Db	2101	ATGTTGGCCACACATATATGATATATCTTGCTGTACTTTAGTATTTTCAACATTTGT	2160
Qy	2161	GAAGAAACATTAAACAAAGTTAAATGTAGTGTGCGAGTTTCTTCCTTCCTTT	2220
Db	2161	GAAGAAACATTAAACAAAGTTAAATGTAGTGTGCGAGTTTCTTCCTTCCTTT	2220
Qy	2221	AAAAATGTTGTTTAAAGCTTTAAACAATGGGAACCCCTTGAGACATGCTCAGATACAT	2280
Db	2221	AAAAATGTTGTTTAAAGCTTTAAACAATGGGAACCCCTTGAGACATGCTCAGATACAT	2280
Qy	2281	GTGGAGAACCAAGAGGCGCTCTTAACTGTAAACAATGTTCATGTTGTGATGTTTTTTTT	2340
Db	2281	GTGGAGAACCAAGAGGCGCTCTTAACTGTAAACAATGTTCATGTTGTGATGTTTTTTTT	2340
Qy	2341	TTTTTTTAAATTAATATTCCAAATGTTTAAATPAAAAAATTTTTTTTTT	2390
Db	2341	TTTTTTTAAATTAATATTCCAAATGTTTAAATPAAAAAATTTTTTTTTT	2390
RESULT 2			
ABK84628			
ID	ABK84628	standard; cDNA; 2390 BP.	
XX	ABK84628;		
AC			
XX			
DT	14-AUG-2002	(first entry)	
DE		Human cDNA differentially expressed in granulocytic cells #1199.	
XX			
KW		Human; ss; granulocytic cell; DNA chip; bacterial infection;	
KW		viral infection; parasitic infection; protozoal infection;	
KW		fungal infection; sterile inflammatory disease; psoriasis;	
KW		rheumatoid arthritis; glomerulonephritis; asthma; thrombosis;	
KW		cardiac reperfusion injury; renal reperfusion injury; AIDS;	
KW		adult respiratory distress syndrome; inflammatory bowel disease;	
KW		Crohn's disease; ulcerative colitis; periodontal disease;	
KW		granulocyte activation; chronic inflammation; allergy.	
XX			
OS	Homo sapiens.		
XX			
PN	WO200228999-A2.		
XX			
PD	11-APR-2002.		
XX			
PF	03-OCT-2001; 2001WO-US030821.		
XX			
PR	03-OCT-2000; 2000US-0237189P.		
XX			
PA	(GENE-) GENE LOGIC INC.		
XX			
FI	Beazer-Barclay Y, Weiseman SM, Yamaga S, Vockley J;		
XX			
DR	WPI; 2002-435328/46.		
XX			

PT	Detecting granulocyte activation by detecting differential expression of genes associated with granulocyte activation, which serves as diagnostic markers that is useful for monitoring disease states and drug toxicity.
PS	Claim 1; SEQ ID NO 1199; 114bp; English.
PX	
XX	The invention relates to detecting (M1) granulocyte (GC) activation (GCA), by detecting the level of expression of gene(s) (Gs) identified by DNA chip analysis as given in the specification, and comparing the expression level to an expression level in an unactivated GC, where differential expression of Gs is indicative of GCA. Also included are modulating (M2) GA by contacting GC with an agent that alters the expression of at least one gene in Gs; (2) screening (M3) for an agent capable of modulating GCA or an inflammation (especially chronic) in a tissue, an allergic response in a subject, exposure of a subject to a pathogen or sterile inflammatory disease using the gene expression profile; (3) detecting (M4) an inflammation (especially chronic) in a tissue, an allergic response in a subject, exposure of a subject to a pathogen or sterile inflammatory disease, by detecting the level of expression in a sample of the tissue of gene(s) from Gs, where the level of expression of the gene is indicative of inflammation; (4) treating (M5) an inflammation (especially chronic) or in a tissue, an allergic response in a subject, exposure of a subject to a pathogen or sterile inflammatory disease, by contacting a tissue having inflammation with an agent that modulates the expression of gene(s) from Gs in the tissue. M1 is useful for detecting GCA; M2 is useful for modulating GA; M3 is useful for screening an agent capable of modulating GCA preferably in an inflammation in a tissue; M4 is useful for detecting an inflammation (especially chronic) in a tissue, an allergic response in a subject, exposure of a subject to a pathogen or sterile inflammatory disease (e.g., psoriasis, rheumatoid arthritis, glomerulonephritis, asthma, thrombosis, cardiac reperfusion injury, renal reperfusion injury, ARDS, adult respiratory distress syndrome, inflammatory bowel disease, Crohn's disease, ulcerative colitis, periodontal disease; also bacterial infection, viral infection, parasitic infection, protozoal infection, fungal infection and MS is useful for treating one of the above conditions. The present sequence represents a gene differentially expressed in granulocytes. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pcc_sequences
SQ	Sequence 2390 BP; 645 A; 589 C; 668 G; 488 T; 0 U; 0 Other;
Query Match	99.9%; Score 2388.4; DB 6; Length 2390;
Best Local Similarity	100.0%; Pred. No. 0;
Matches 2389; Conservative	0; Mismatches 1; Indels 0; Gaps 0;
OY	1 AGAGGGAGACGGACGTTGAGGAACGAGGAGAAAGAGAAGAAAATGCGTCCACGGATTAA 60
DB	1 AGAGGGAGACGGAGGTGAGGAACGAGGAGAAAGAGAAGAAAATGCGTCCACGGATTAA 60
OY	CAGTACCCTAATGCCAAGCTGACGGGGCAGGAGGCTCACAGTGCTTAACACCGGACGCCAC 120
DB	61 CAGTACCCTAATGCCAAGCTGACGGGGCAGGAGGCTCACAGTGCTTAACACCGGACGCCAC 120
OY	121 TCAAAGATATGACAGAACCAACCGACCATATGAGCAACAAAGCTATGGAACCTATGACA 180
DB	121 TCAAAGATATGACAGAACCAACCGACCATATGAGCAACAAAGCTATGGAACCTATGACA 180
OY	181 GCCCATGATATGACGCTTAATCCCAAGGCTCACAACAATGCAACCTATATGGGACAGACGCTTA 240
DB	181 GCCCATGATATGACGCTTAATCCCAAGGCTCACAACAATGCAACCTATATGGGACAGACGCTTA 240
OY	241 TGCAACTTCCTATGAGACAGCTCCACATGCTTAACTACTCAATCGCCCCCAGGACATA 300
DB	241 TGCAACTTCCTATGAGACAGCTCCACATGCTTAACTACTCAATCGCCCCCAGGACATA 300
OY	301 CAGCCAGCTCTTCCAGGGGATATGGCACTGATGCTTATGATACACACATGCTTAAGTAC 360
DB	301 CAGCCAGCTCTTCCAGGGGATATGGCACTGATGCTTATGATACACACATGCTTAAGTAC 360
OY	361 CAGCACCAGGAGCTCCCTATGAGAGCTCAAGTCTGATATATGAGGACCTAGAGCTGCTTATCCAGC 420

Db 361 CACACCCAGGCTCTTATGCACTAGTCTGATATGGCACTGACCTGCTTATTCAC 420  
 Qy 421 CTATGGGAGCAGCAGCAGCAGCCTGACCTTACAGACCGAGATGAAACAGCCAC 480  
 Db 421 CTATGGGAGCAGCAGCAGCAGCCTGACCTTACAGACCGAGATGAAACAGCCAC 480  
 Qy 481 TGAAGCTTACTCAACCTCAATTTAGCAAGGGGTTTCAACAGCCAGCTGAGATATG 540  
 Db 481 TGAAGCTTACTCAACCTCAATTTAGCAAGGGGTTTCAACAGCCAGCTGAGATATG 540  
 Qy 541 ACAGAGTAACCTCAAGTTATCCAGGAGTACCTGAGAGTACCCCATGACAGCTAC 600  
 Db 541 ACAGAGTAACCTCAAGTTATCCAGGAGTACCTGAGAGTACCCCATGACAGCTAC 600  
 Qy 601 ACCTCATCTTACCTCTTACAGCTATTCCTTACAGCCAGCTAGTTATGATGAG 660  
 Db 601 ACCTCATCTTACCTCTTACAGCTATTCCTTACAGCCAGCTAGTTATGATGAG 660  
 Qy 661 CAGTTACTGTCAGAGAACCTTATGGGAGACCGAGAGCTATGACAGAGTACGTA 720  
 Db 661 CAGTTACTGTCAGAGAACCTTATGGGAGACCGAGAGCTATGACAGAGTACGTA 720  
 Qy 721 TGGTCAACAAAGAGCTATGGGAGAGCTCCCACTAGTTACCCACCCCAACTGAT 780  
 Db 721 TGGTCAACAAAGAGCTATGGGAGAGCTCCCACTAGTTACCCACCCCAACTGAT 780  
 Qy 781 CTACAGCCAGCTCTCAAGTCAATATAGCCACAGACAGAGCTAGCGGAGAGATTC 840  
 Db 781 CTACAGCCAGCTCTCAAGTCAATATAGCCACAGACAGAGCTAGCGGAGAGATTC 840  
 Qy 841 ATTCCCAAGAGACCCAGTACAGGAGTATGGGAGAGCTAGGAGATTTTC 900  
 Db 841 ATTCCCAAGAGACCCAGTACAGGAGTATGGGAGAGCTAGGAGATTTTC 900  
 Qy 901 CGGACAGAGAGACCCAGGAGTACAGGAGTATGGGAGAGCTAGGAGATTTTC 960  
 Db 901 CGGACAGAGAGACCCAGGAGTACAGGAGTATGGGAGAGCTAGGAGATTTTC 960  
 Qy 961 ATTGATCGTGAAGAGTACAGGAGTATGGGAGAGCTAGGAGATTTTC 1020  
 Db 961 ATTGATCGTGAAGAGTACAGGAGTATGGGAGAGCTAGGAGATTTTC 1020  
 Qy 1021 CGTGAAGAGAGGAGTATTAAGAGCTGAGAGCCATGAGTGAAGAGAGATTC 1080  
 Db 1021 CGTGAAGAGAGGAGTATTAAGAGCTGAGAGCCATGAGTGAAGAGAGATTC 1080  
 Qy 1081 TGAATCAGGAGAGCTGAGAGTACAGTGAAGAGCTGAGAGAGTATTTATGAC 1140  
 Db 1081 TGAATCAGGAGAGCTGAGAGTACAGTGAAGAGCTGAGAGAGTATTTATGAC 1140  
 Qy 1141 AGGATTAATGACAGTGTGACTTATGATCTGAGAGATTTTAAAGAGTGGGT 1200  
 Db 1141 AGGATTAATGACAGTGTGACTTATGATCTGAGAGATTTTAAAGAGTGGGT 1200  
 Qy 1201 TGTTAAGATGACAGAGAGTGGGAGAGCTGATGATGATGATGATGATGATGAT 1260  
 Db 1201 TGTTAAGATGACAGAGAGTGGGAGAGCTGATGATGATGATGATGATGATGAT 1260  
 Qy 1261 AGGAAAGCCCAAGAGAGTGCACAGTGTCTTATGAAGAGCCAGCTGCAAGGTGC 1320  
 Db 1261 AGGAAAGCCCAAGAGAGTGCACAGTGTCTTATGAAGAGCCAGCTGCAAGGTGC 1320  
 Qy 1321 CGTGAAGTGTGATGAGAGAGTATTTCAAGGAGAGAGTATTTTCTGCTGCTG 1380  
 Db 1321 CGTGAAGTGTGATGAGAGAGTATTTCAAGGAGAGAGTATTTTCTGCTGCTG 1380  
 Qy 1381 GAGAGAGCTCTCAATGAGAGTATGAGGAGTATGAGGAGTATGAGGAGTATGAG 1440  
 Db 1381 GAGAGAGCTCTCAATGAGAGTATGAGGAGTATGAGGAGTATGAGGAGTATGAG 1440  
 Qy 1441 GCCACCAACACTCGTGAAGTCCAGAGAGCCAGAGAGTCTGAGGAGTCCATGAGTGC 1500  
 Db 1441 GCCACCAACACTCGTGAAGTCCAGAGAGCCAGAGAGTCTGAGGAGTCCATGAGTGC 1500

Db 1441 GCCACCAACACTCGTGAAGTCCAGAGAGCCAGAGAGTCTGAGGAGTCCATGAGTGC 1500  
 Qy 1501 CATGGAGAGCTGTGAGAGATGAGAGGCTTCCCTCAAGAGAGACCCGGGGTTCCCG 1560  
 Db 1501 CATGGAGAGCTGTGAGAGATGAGAGGCTTCCCTCAAGAGAGACCCGGGGTTCCCG 1560  
 Qy 1561 AGGAAAGCTCTGAGAGAGAGAGTCCAGAGAGCTGAGAGTCCAGTGTCCAA 1620  
 Db 1561 AGGAAAGCTCTGAGAGAGAGAGTCCAGAGAGCTGAGAGTCCAGTGTCCAA 1620  
 Qy 1621 TCCGGGTTGTGAACCAAGACTTCCCTGAGAGAGAGTCCAGTGTCCAA 1680  
 Db 1621 TCCGGGTTGTGAACCAAGACTTCCCTGAGAGAGAGTCCAGTGTCCAA 1680  
 Qy 1681 AAGGCTGAGAGCTTCCCTGAGAGAGAGTCCCTGAGAGAGTCCCTGAGAGAG 1740  
 Db 1681 AAGGCTGAGAGCTTCCCTGAGAGAGAGTCCCTGAGAGAGTCCCTGAGAGAG 1740  
 Qy 1741 TGGCCCTGAGAGTCCCTGAGAGAGAGTCCCTGAGAGAGTCCCTGAGAGAG 1800  
 Db 1741 TGGCCCTGAGAGTCCCTGAGAGAGAGTCCCTGAGAGAGTCCCTGAGAGAG 1800  
 Qy 1801 AATGTTCAAGAGTCCCTGAGAGAGAGTCCCTGAGAGAGTCCCTGAGAGAG 1860  
 Db 1801 AATGTTCAAGAGTCCCTGAGAGAGAGTCCCTGAGAGAGTCCCTGAGAGAG 1860  
 Qy 1861 CGAGGAGTCCCTGAGAGAGAGTCCCTGAGAGAGTCCCTGAGAGAGTCCCTGAG 1920  
 Db 1861 CGAGGAGTCCCTGAGAGAGAGTCCCTGAGAGAGTCCCTGAGAGAGTCCCTGAG 1920  
 Qy 1921 GGAACAGATGAGAGAGAGAGAGTCCCTGAGAGAGTCCCTGAGAGAGTCCCTGAG 1980  
 Db 1921 GGAACAGATGAGAGAGAGAGAGTCCCTGAGAGAGTCCCTGAGAGAGTCCCTGAG 1980  
 Qy 1981 GCAAGTCCAGAGAGAGAGAGTCCCTGAGAGAGTCCCTGAGAGAGTCCCTGAG 2040  
 Db 1981 GCAAGTCCAGAGAGAGAGAGTCCCTGAGAGAGTCCCTGAGAGAGTCCCTGAG 2040  
 Qy 2041 GACTACAGATTTATTTTAAACAGAGAGTCCCTGAGAGAGTCCCTGAGAGAG 2100  
 Db 2041 GACTACAGATTTATTTTAAACAGAGAGTCCCTGAGAGAGTCCCTGAGAGAG 2100  
 Qy 2101 ATGTTGAGAGAGAGTCCCTGAGAGAGTCCCTGAGAGAGTCCCTGAGAGAG 2160  
 Db 2101 ATGTTGAGAGAGAGTCCCTGAGAGAGTCCCTGAGAGAGTCCCTGAGAGAG 2160  
 Qy 2161 GAGAGAGCTTTAAACAGAGTCCCTGAGAGAGTCCCTGAGAGAGTCCCTGAG 2220  
 Db 2161 GAGAGAGCTTTAAACAGAGTCCCTGAGAGAGTCCCTGAGAGAGTCCCTGAG 2220  
 Qy 2221 AAAAATGTTTAAAGTCCCTGAGAGAGTCCCTGAGAGAGTCCCTGAGAGAG 2280  
 Db 2221 AAAAATGTTTAAAGTCCCTGAGAGAGTCCCTGAGAGAGTCCCTGAGAGAG 2280  
 Qy 2281 GTGAGAGAGAGAGAGTCCCTGAGAGAGTCCCTGAGAGAGTCCCTGAGAGAG 2340  
 Db 2281 GTGAGAGAGAGAGAGTCCCTGAGAGAGTCCCTGAGAGAGTCCCTGAGAGAG 2340  
 Qy 2341 TTTTAAATTAATTTCAATGTTTAAATTAATTAATTAATTAATTAATTAATTA 2390  
 Db 2341 TTTTAAATTAATTTCAATGTTTAAATTAATTAATTAATTAATTAATTAATTA 2390  
 RESULT 3  
 ABK64822  
 ID ABK64822 standard; DNA; 2390 BP.  
 XX  
 AC ABK64822;  
 XX  
 DE 18-JUN-2002 (first entry)  
 XX Human benign prostatic hyperplasia gene #717.  
 XX

KM Human; benign prostatic hyperplasia; BPH; prostate cancer; gene; ds.  
 XX Homo sapiens.  
 OS WO000212440-A2.  
 XX 14-FEB-2002.  
 PD 07-AUG-2001, 2001MO-US024708.  
 XX 07-AUG-2000, 2000US-0223323P.  
 PR 05-JUN-2001, 2001US-00873319.  
 XX (GENE-) GENE LOGIC INC.  
 PA (NIBS) JAPAN TOBACCO INC.  
 XX Munger WE, Kulkarini P, Getzenberg RH, Waga I, Yamamoto J;  
 PI WPI, 2002-257476/30.  
 DR  
 XX  
 XX  
 PT Identifying drugs for and diagnosing benign prostatic hyperplasia, by  
 PT detecting expression levels of one or more genes in prostate cells from  
 PT patient that are differentially regulated compared to normal prostate  
 PT cells.  
 PS Disclosure; Page 399-400; 444pp; English.  
 CC The invention relates to a method of diagnosing (I) the onset or  
 CC progression of benign prostatic hyperplasia (BPH), or screening (II) for  
 CC or identifying an agent that modulates the onset or progression of BPH.  
 CC The method is based on changes in gene expression in BPH tissue isolated  
 CC from patients exhibiting different clinical states of prostate  
 CC hyperplasia as compared to normal prostate tissue. (I) comprises  
 CC detecting the expression levels of one or more genes in prostate cells  
 CC from the subject that are differentially regulated compared to normal  
 CC prostate cells. (II) comprises preparing a first gene expression profile  
 CC of BPH cells or BPH-like cell population, exposing the cells to the  
 CC agent, preparing a second gene expression profile of the agent exposed  
 CC cells, and comparing the first and second gene expression profiles. (I)  
 CC is useful for diagnosing the onset or progression of BPH. (II) is useful  
 CC for identifying an agent that modulates the onset or progression of BPH.  
 CC The methods are useful to present information identifying the expression  
 CC level in a tissue or cells, by comparing the expression level of genes  
 CC given in the specification in the tissue or cells to the level of  
 CC expression of gene in the database, and displaying the expression levels  
 CC of at least one gene in the tissue or cell sample compared to the  
 CC expression level in BPH. Agents using (II) are useful for treating BPH or  
 CC prostate cancer. ABK64106-ABK64860 represent human benign prostatic  
 CC hyperplasia gene sequences of the invention  
 CC  
 XX  
 SQ Sequence 2390 BP, 645 A, 589 C, 668 G, 488 T, 0 U, 0 Other;  
 Query Match 99.9%; Score 2388.4; DB 6; Length 2390;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 2389; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 241 TGCACCTTCTTATGACAGCCTCCACTGTTATCTACTCCAACTGCCCCCAGGAGATA 300  
 QY 301 CAGCCAGCCTGTCCAGGGGTATGGCACTGGTCTTATGATACCACTGCTCACTGCTAC 360  
 Db 301 CAGCCAGCCTGTCCAGGGGTATGGCACTGGTCTTATGATACCACTGCTCACTGCTAC 360  
 QY 361 CACACCCAGGCTCTTATGACAGCTCACTGCTATGAGCACTGAGCTGCTTATCCAGC 420  
 Db 361 CACACCCAGGCTCTTATGACAGCTCACTGCTATGAGCACTGAGCTGCTTATCCAGC 420  
 QY 421 CTATGGGAGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 480  
 Db 421 CTATGGGAGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 480  
 QY 481 TGAGACTGATCAACTCTTAACTTACAGAGGGGTTCACACAGCCAGCCTAGATATGG 540  
 Db 481 TGAGACTGATCAACTCTTAACTTACAGAGGGGTTCACACAGCCAGCCTAGATATGG 540  
 QY 541 ACAGAGTAACTACAGTTATCCCAAGTACCTGGAGCTACCCATGACAGCTCACTG 600  
 Db 541 ACAGAGTAACTACAGTTATCCCAAGTACCTGGAGCTACCCATGACAGCTCACTG 600  
 QY 601 ACCTCATCTCTCCCTCCTACAGAGCTATTCCTTACAGAGCCGACTAGTTATGATGAG 660  
 Db 601 ACCTCATCTCTCCCTCCTACAGAGCTATTCCTTACAGAGCCGACTAGTTATGATGAG 660  
 QY 661 CAGTTACTCTCAGCAGAAACCTATGAGCAACGAGCAGCTATGACAGAGTATGACTA 720  
 Db 661 CAGTTACTCTCAGCAGAAACCTATGAGCAACGAGCAGCTATGACAGAGTATGACTA 720  
 QY 721 TGGTCAACAAAGAGCTATGGGAGAGCCTCCACTAGTTACCCACCCCAACTGATC 780  
 Db 721 TGGTCAACAAAGAGCTATGGGAGAGCCTCCACTAGTTACCCACCCCAACTGATC 780  
 QY 781 CTACAGCCAGGCTCAAGTCAATATATGCAACAGAGCAGAGCTACGGGAGAGAGTTC 840  
 Db 781 CTACAGCCAGGCTCAAGTCAATATATGCAACAGAGCAGAGCTACGGGAGAGAGTTC 840  
 QY 841 ATTCCGACAGAGCAACCCAGTATGAGTGGTCTTATGAGGAGAGTCTGAGAGATTTTC 900  
 Db 841 ATTCCGACAGAGCAACCCAGTATGAGTGGTCTTATGAGGAGAGTCTGAGAGATTTTC 900  
 QY 901 CGAGCCAGAGAAACCGAGCATGATGAGCCTGATTAACCGGGGAGGGGAGAGGGG 960  
 Db 901 CGAGCCAGAGAAACCGAGCATGATGAGCCTGATTAACCGGGGAGGGGAGAGGGG 960  
 QY 961 ATTGATGCTGAGAGATGAGAGAGAGTGGGGGAGAGAGAGAGAGAGAGAGAGAGAG 1020  
 Db 961 ATTGATGCTGAGAGATGAGAGAGAGTGGGGGAGAGAGAGAGAGAGAGAGAGAGAG 1020  
 QY 1021 CGCTGAGAGAGAGAGTGGCTTCAATATAGCCTGTGAGCCATGATGAGAGCAAGATCT 1080  
 Db 1021 CGCTGAGAGAGAGAGTGGCTTCAATATAGCCTGTGAGCCATGATGAGAGCAAGATCT 1080  
 QY 1081 TGATCTAGGCTCTCTGTATGATCCAGATGAAGACTCTGACAACTGACAACTTATTA 1140  
 Db 1081 TGATCTAGGCTCTCTGTATGATCCAGATGAAGACTCTGACAACTGACAACTTATTA 1140  
 QY 1141 AGGATTAATGACATGATGACCTCTGATGATATGAGCAACTCTTAAAGATGAGG 1200  
 Db 1141 AGGATTAATGACATGATGACCTCTGATGATATGAGCAACTCTTAAAGATGAGG 1200  
 QY 1201 TGTTAAGATGAACAAGAGATGAGCAACCTATGATCAATCTACCTGAGCAAGAAAC 1260  
 Db 1201 TGTTAAGATGAACAAGAGATGAGCAACCTATGATCAATCTACCTGAGCAAGAAAC 1260  
 QY 1261 AGGAAAGCCCAAGAGAGATGAGCAAGTGTCTTATGAGAGAGAGAGAGAGAGAGAG 1320  
 Db 1261 AGGAAAGCCCAAGAGAGATGAGCAAGTGTCTTATGAGAGAGAGAGAGAGAGAGAG 1320  
 QY 1321 CGTGAAGTGTGATGAGAGAGAGATTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1380

```

Db 1321 CGGGAATGTTTGTATGGAGAAATTTTCAAGGAGCAAACTTAAAGTCTCCCTTGCTCG 1380
Qy 1381 GAAGAAAGCTCCCAATGAACATATGCGGGGTGTCTGCGCAACCCCGTGAAGGCGAGAGGCAT 1440
Db 1381 GAAGAAAGCTCCCAATGAACATATGCGGGGTGTCTGCGCAACCCCGTGAAGGCGAGAGGCAT 1440
Qy 1441 GCCACCAACATCTCCGTGAGAGTCCAGAGAGCCAGAGAGGCTCTGAGGAGCCCATGAGGTG 1500
Db 1441 GCCACCAACATCTCCGTGAGAGTCCAGAGAGCCAGAGAGGCTCTGAGGAGCCCATGAGGTG 1500
Qy 1501 CATGGAGAGCCGTGAGAGATAGAGAGGCTTCCCTCCAGAGAGCCCGAGGCTTCCCG 1560
Db 1501 CATGGAGAGCCGTGAGAGATAGAGAGGCTTCCCTCCAGAGAGCCCGAGGCTTCCCG 1560
Qy 1561 AGGGAACCCCTCTGAGAGAGAAAGCTCCAGACCCAGAGTGAAGATGGCAATGTCCCA 1620
Db 1561 AGGGAACCCCTCTGAGAGAGAAAGCTCCAGACCCAGAGTGAAGATGGCAATGTCCCA 1620
Qy 1621 TCCGGGTTGTGAAACAGAACTTCGCTGAGAGAGAGAGTGCACAGTGAAGGCCCC 1680
Db 1621 TCCGGGTTGTGAAACAGAACTTCGCTGAGAGAGAGAGTGCACAGTGAAGGCCCC 1680
Qy 1681 AAAGCCTGAAGGCTTCTCCCGCACCTTTCGCCCCCGAGGTGTGATGTGAGAGAG 1740
Db 1681 AAAGCCTGAAGGCTTCTCCCGCACCTTTCGCCCCCGAGGTGTGATGTGAGAGAG 1740
Qy 1741 TGGCCCTGTGTGATGCGGGAGAGAGAGAGTGGCTCTATGATCTGTGTGTCTCCGCTG 1800
Db 1741 TGGCCCTGTGTGATGCGGGAGAGAGAGAGTGGCTCTATGATCTGTGTGTCTCCGCTG 1800
Qy 1801 AATGTTCAAGAGTGTGCGGTGTGAGAGAGAGTGGCTTCCGCTGTGCGGGGAGTGA 1860
Db 1801 AATGTTCAAGAGTGTGCGGTGTGAGAGAGAGTGGCTTCCGCTGTGCGGGGAGTGA 1860
Qy 1861 CCGAGGTGCTTGTGTGAGAGAGAGAGTGGCTTGTGAGAGGAGCCCTGTGAGCTTTGAT 1920
Db 1861 CCGAGGTGCTTGTGTGAGAGAGAGAGTGGCTTGTGAGAGGAGCCCTGTGAGCTTTGAT 1920
Qy 1921 GGAACGATGTGGAGAGAGAGAGAGAGAGTGGAGAGCTGTGAGAGAGAGTGAAGGCA 1980
Db 1921 GGAACGATGTGGAGAGAGAGAGAGAGAGTGGAGAGAGCTGTGAGAGAGAGTGAAGGCA 1980
Qy 1981 GCACGCTCAGAGAGAGAGATGCGGCTCTAGATGAGAGAGAGAGAGAGAGTGAAGGCA 2040
Db 1981 GCACGCTCAGAGAGAGAGATGCGGCTCTAGATGAGAGAGAGAGAGAGAGTGAAGGCA 2040
Qy 2041 GACTACCAATTTATTTTAAACAGAAATGTTTTAAATTTAAATTCATATTTATA 2100
Db 2041 GACTACCAATTTATTTTAAACAGAAATGTTTTAAATTTAAATTCATATTTATA 2100
Qy 2101 AATGTTGGCCCAACATTATGATATTCCTGTGATCTTATCTTATTTTACCATTTGT 2160
Db 2101 AATGTTGGCCCAACATTATGATATTCCTGTGATCTTATCTTATTTTACCATTTGT 2160
Qy 2161 GAAGAAACATTAAACAAGTTAATGTAATGTCGAGATTTTCTTCTCTTCTTCTT 2220
Db 2161 GAAGAAACATTAAACAAGTTAATGTAATGTCGAGATTTTCTTCTCTTCTTCTTCTT 2220
Qy 2221 AAAAATGTTTAAAGATTTTAAACATGAGAAACCCCTGTGAGAGAGTCTCAATTCAT 2280
Db 2221 AAAAATGTTTAAAGATTTTAAACATGAGAAACCCCTGTGAGAGAGTCTCAATTCAT 2280
Qy 2281 GTGGAAGAACCAAGAGGCTCTTAACTGTAACAATGTTCATGTGTGTGATGTTTCTT 2340
Db 2281 GTGGAAGAACCAAGAGGCTCTTAACTGTAACAATGTTCATGTGTGTGATGTTTCTT 2340
Qy 2341 TTTTAAAAATTAATTCCTCAATGTTTAAATTAATTAATTAATTAATTAATTAATTA 2390
Db 2341 TTTTAAAAATTAATTCCTCAATGTTTAAATTAATTAATTAATTAATTAATTAATTA 2390

```

RESULT 4  
ABN97274

```

ID ABN97274 standard; DNA; 2390 BP.
XX
AC ABN97274;
XX
DT 13-AUG-2002 (first entry)
XX
DE Gene #3772 used to diagnose liver cancer.
XX
KW Gene; liver cancer; ds; hepatocellular carcinoma; hepatotropic;
XX metastatic liver tumour; cytostatic; expression profile; disease state;
XX disease progression; drug toxicity; drug efficacy; drug metabolism.
XX
OS Homo sapiens.
XX
PN MO200229103-A2.
XX
PD 11-APR-2002.
XX
PF 02-OCT-2001; 2001WO-05030589.
XX
PR 02-OCT-2000; 2000US-0237054P.
XX
PA (GENE-) GENE LOGIC INC.
XX
PI Horne D, Alvares C, Peres-Da-Silva S, Vockley JG;
XX
PI WPI; 2002-426119/45.
XX
DR Diagnosing and detecting the progression of liver cancer, hepatocellular
XX carcinoma or metastatic liver tumor in a patient, involves detecting the
XX level of expression of two or more genes in a liver tissue sample.
XX
PS Claim 1: SEQ ID NO 3772; 298bp; English.
XX
CC The invention relates to a novel method for diagnosing and detecting the
XX progression of liver cancer, hepatocellular carcinoma or metastatic liver
XX tumor in a patient, and differentiating metastatic liver cancer from
XX hepatocellular carcinoma in a patient, involving detecting the level of
XX expression of two or more genes represented in ABN93503-ABN97455 in a
XX tissue sample. The method of the invention has hepatotropic, and
XX cytostatic activity. The method is useful for diagnosing and detecting
XX the progression of liver cancer, hepatocellular carcinoma and metastatic
XX liver carcinoma in a patient. The method is useful for identifying
XX expression profiles which serve as useful diagnostic markers as well as
XX markers that can be used to monitor disease states, disease progression,
XX drug toxicity, drug efficacy and drug metabolism. Note: The sequence data
XX for this patent did not form part of the printed specification, but was
XX obtained in electronic form directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 2390 BP; 645 A; 589 C; 668 G; 488 T; 0 U; 0 Other;
XX
Query Match 99.9%; Score 2388.4; DB 6; Length 2390;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2389; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 1 AGAGGAGAGCGGACCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGATTA 60
Db 1 AGAGGAGAGCGGACCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGATTA 60
Qy 61 CAGTACCTTATGACCAAGCTGACAGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGATTA 120
Db 61 CAGTACCTTATGACCAAGCTGACAGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGATTA 120
Qy 121 TCAAGGATATGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGATTA 180
Db 121 TCAAGGATATGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGATTA 180
Qy 181 GCCCACTGATGTGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGATTA 240
Db 181 GCCCACTGATGTGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGATTA 240
Qy 241 TGAACCTTATGAGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGATTA 300

```



Db 241 TGAACCTTCTTATGAGACGCTCCCACTGGTTATCTCAACTGCCCCCGAGGATTA 300  
Qy 301 CAGCCAGCCTGTCAGAGGGATATGCACTGCTGCTTATGATACCACTGCTCACTAGTAC 360  
Db 301 CAGCCAGCCTGTCAGAGGGATATGCACTGCTGCTTATGATACCACTGCTCACTAGTAC 360  
Qy 361 CAGCCAGCCTGTCAGAGGGATATGCACTGCTGCTTATGATACCACTGCTCACTAGTAC 420  
Db 361 CAGCCAGCCTGTCAGAGGGATATGCACTGCTGCTTATGATACCACTGCTCACTAGTAC 420  
Qy 421 CTATGGGACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 480  
Db 421 CTATGGGACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 480  
Qy 481 TGAAGTACTGCACTCTCAATCTAGCAAGGGGGTTTACACAGCCAGCCTAGATATGG 540  
Db 481 TGAAGTACTGCACTCTCAATCTAGCAAGGGGGTTTACACAGCCAGCCTAGATATGG 540  
Qy 541 ACAGAGTAACTAGAGTTATCCCAAGTACCTGGAGAGTACCCCAATGAGCAGTCACTGC 600  
Db 541 ACAGAGTAACTAGAGTTATCCCAAGTACCTGGAGAGTACCCCAATGAGCAGTCACTGC 600  
Qy 601 ACCTCACTCTACCTCTCTACAGCTATCTCTCTACAGCAGCAGCAGCAGCAGCAGCAG 660  
Db 601 ACCTCACTCTACCTCTCTACAGCTATCTCTCTACAGCAGCAGCAGCAGCAGCAGCAG 660  
Qy 661 CAGTTACTCTCAGCAGAACCTATGGGCAACGAGCAGCTATGAGCAGAGTACTA 720  
Db 661 CAGTTACTCTCAGCAGAACCTATGGGCAACGAGCAGCTATGAGCAGAGTACTA 720  
Qy 721 TGGTCAACAAAGAGCTATGGGCAAGCCTCCCACTAGTTACCCCACTGATC 780  
Db 721 TGGTCAACAAAGAGCTATGGGCAAGCCTCCCACTAGTTACCCCACTGATC 780  
Qy 781 CTACAGCAGAGCTCAAGTCAATATAGCAAGAGCAGCAGCAGCAGCAGCAGCAGCAG 840  
Db 781 CTACAGCAGAGCTCAAGTCAATATAGCAAGAGCAGCAGCAGCAGCAGCAGCAGCAG 840  
Qy 841 ATTCCGACAGAGCAACCCAGTAGAGTGGTGTATATGGCAGAGAGTCTGAGAGTTTTC 900  
Db 841 ATTCCGACAGAGCAACCCAGTAGAGTGGTGTATATGGCAGAGAGTCTGAGAGTTTTC 900  
Qy 901 CGGACAGAGAGAACCGGAGCATGAGTGGCCTGATTAACGGGGGAGGGAGAGGGGG 960  
Db 901 CGGACAGAGAGAACCGGAGCATGAGTGGCCTGATTAACGGGGGAGGGAGAGGGGG 960  
Qy 961 ATTGATCGTGGAGCATGAGAGTGGGCGGGGAGAGAGCGGGTGAATGGGCGAG 1020  
Db 961 ATTGATCGTGGAGCATGAGAGTGGGCGGGGAGAGAGCGGGTGAATGGGCGAG 1020  
Qy 1021 CGCTGGAGAGCGAGTGGCTTCAATAAGCCTGGTGGACCCATGATGAGAGCAAGATCT 1080  
Db 1021 CGCTGGAGAGCGAGTGGCTTCAATAAGCCTGGTGGACCCATGATGAGAGCAAGATCT 1080  
Qy 1081 TGAATCTAGGCGCCCTCTGATGATCAAGTGAAGACTGACACAGAGTCAATTTATGACA 1140  
Db 1081 TGAATCTAGGCGCCCTCTGATGATCAAGTGAAGACTGACACAGAGTCAATTTATGACA 1140  
Qy 1141 AGGATTAATGACAGTGTGACTAGTATCTGGAGACTTCTTTAAGAGTGGGGGT 1200  
Db 1141 AGGATTAATGACAGTGTGACTAGTATCTGGAGACTTCTTTAAGAGTGGGGGT 1200  
Qy 1201 TGTTAAGATGAAACAAGAACTGGGCAACCATGATCCATCTACCTGGACAAGAAAC 1260  
Db 1201 TGTTAAGATGAAACAAGAACTGGGCAACCATGATCCATCTACCTGGACAAGAAAC 1260  
Qy 1261 AGGAAAGCCCAAGGGGATGCCAGTGTCTATGAAGACCCCACTGCGCAAGGCTGC 1320  
Db 1261 AGGAAAGCCCAAGGGGATGCCAGTGTCTATGAAGACCCCACTGCGCAAGGCTGC 1320  
Qy 1321 CGTGAATGGTTGATGGGAAAGATTTCAAGGAGCAACTTAAAGTCCCTGCTGC 1380  
Db 1321 CGTGAATGGTTGATGGGAAAGATTTCAAGGAGCAACTTAAAGTCCCTGCTGC 1380

Db 1321 CGTGAATGGTTGATGGGAAAGATTTCAAGGAGCAACTTAAAGTCCCTGCTGC 1380  
Qy 1381 GAAAGAGCTCCAAATGACATATGCGGGTGGTCTGCAACCCCGTGAAGGAGAGGAT 1440  
Db 1381 GAAAGAGCTCCAAATGACATATGCGGGTGGTCTGCAACCCCGTGAAGGAGAGGAT 1440  
Qy 1441 GCAACACCACTCCCGGAGAGTCCAGAGAGCCCAAGAGTCCCGGAGAGCCCAATGGGTCG 1500  
Db 1441 GCAACACCACTCCCGGAGAGTCCAGAGAGCCCAAGAGTCCCGGAGAGCCCAATGGGTCG 1500  
Qy 1501 CATGGAGAGCCCGTGAAGAGATAGAGAGGCTTCCCTCCAAAGAGAGCCCGGGGTTCCCG 1560  
Db 1501 CATGGAGAGCCCGTGAAGAGATAGAGAGGCTTCCCTCCAAAGAGAGCCCGGGGTTCCCG 1560  
Qy 1561 AGGAAACCCCTCTGAGAGAGAAACCTCAGACCCAGCTGAGAGCTGGCAGTGTCCCA 1620  
Db 1561 AGGAAACCCCTCTGAGAGAGAAACCTCAGACCCAGCTGAGAGCTGGCAGTGTCCCA 1620  
Qy 1621 TCCGGGTTGGTGAACCAAGACTTCCCTGGAGAAACAGAGTCAACAGTGAAGGCCCC 1680  
Db 1621 TCCGGGTTGGTGAACCAAGACTTCCCTGGAGAAACAGAGTCAACAGTGAAGGCCCC 1680  
Qy 1681 AAAGCTGAAGGCTTCTCCCGCACTTCCCGCCCGGGTGGTGAATGCTGGCAGAGG 1740  
Db 1681 AAAGCTGAAGGCTTCTCCCGCACTTCCCGCCCGGGTGGTGAATGCTGGCAGAGG 1740  
Qy 1741 TGGCCCTGGTGCATGCGGAGAGAAAGAGTGGCTCATGATGCTGGTGGTCCGGTGG 1800  
Db 1741 TGGCCCTGGTGCATGCGGAGAGAAAGAGTGGCTCATGATGCTGGTGGTCCGGTGG 1800  
Qy 1801 AATGTTCAAGAGTGGGCGCTGCTGCTGAGACAGAGTGGCTTCCGTGGTGGCGGGCATGA 1860  
Db 1801 AATGTTCAAGAGTGGGCGCTGCTGCTGAGACAGAGTGGCTTCCGTGGTGGCGGGCATGA 1860  
Qy 1861 CCGAGGTGGCTTGGTGGAGAGAAAGAGTGGCCCTGGGGGGGCCCTTGACCTTTGAT 1920  
Db 1861 CCGAGGTGGCTTGGTGGAGAGAAAGAGTGGCCCTGGGGGGGCCCTTGACCTTTGAT 1920  
Qy 1921 GGAACAGATGGAGAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1980  
Db 1921 GGAACAGATGGAGAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1980  
Qy 1981 GCAACCTCAGAGAGCGAGAGATGGGCTTCTAGATGAGAGAGAGAGAGAGAGAGAG 2040  
Db 1981 GCAACCTCAGAGAGCGAGAGATGGGCTTCTAGATGAGAGAGAGAGAGAGAGAGAG 2040  
Qy 2041 GACTACAGATTTATTTTAAACAGAAATGTTTAAATTTAATTTCAATTTATTA 2100  
Db 2041 GACTACAGATTTATTTTAAACAGAAATGTTTAAATTTAATTTCAATTTATTA 2100  
Qy 2101 ATGTTGGCACAACATTAATTAATCTTGTCTGTACTTATGATTTTTCACATTTGT 2160  
Db 2101 ATGTTGGCACAACATTAATTAATCTTGTCTGTACTTATGATTTTTCACATTTGT 2160  
Qy 2161 GAAAGAACTTAAACAGTAAATGATGAGTGGAGGATTTTCTCTCTCTT 2220  
Db 2161 GAAAGAACTTAAACAGTAAATGATGAGTGGAGGATTTTCTCTCTCTT 2220  
Qy 2221 AAAAATGTTGTTTAAAGCTTTAAACAAAGGAGAGAGAGAGAGAGAGAGAGAGAG 2280  
Db 2221 AAAAATGTTGTTTAAAGCTTTAAACAAAGGAGAGAGAGAGAGAGAGAGAGAGAG 2280  
Qy 2281 GTGAGAACCAAGAGGCTCTTAACTGTAACTGTAACATGTTTCAATGTTTGT 2340  
Db 2281 GTGAGAACCAAGAGGCTCTTAACTGTAACTGTAACATGTTTCAATGTTTGT 2340  
Qy 2341 TTTTAAATTAATTAATTTCAATGTTTAAATTAATTAATTAATTAATTAATTAATTA 2390  
Db 2341 TTTTAAATTAATTAATTTCAATGTTTAAATTAATTAATTAATTAATTAATTAATTA 2390  
RESULT 5  
ACN40903



```

QY 1201 TGTAAAGTGAAGAGAACTGGGCAACCATGATCCACATCTACTGAGCAAGAAAC 1260
DB 1201 TGTAAAGTGAAGAGAACTGGGCAACCATGATCCACATCTACTGAGCAAGAAAC 1260
QY 1261 AGGAAAGCCCAAGAGCGATGCCACAGTGTCTTATGAAGACCAACCACTGCCAAGGCTGC 1320
DB 1261 AGGAAAGCCCAAGAGCGATGCCACAGTGTCTTATGAAGACCAACCACTGCCAAGGCTGC 1320
QY 1321 CGTGAATGTTGATGGGAAAGATTTTCAAGGAGCAAACTTAAAGTCTCCCTTGCTCG 1380
DB 1321 CGTGAATGTTGATGGGAAAGATTTTCAAGGAGCAAACTTAAAGTCTCCCTTGCTCG 1380
QY 1381 GAAGAAGCTCTCAATGAACAGTATGCGGGGTGTCTGCCACCCCTGAGAGGCGAGAGCAT 1440
DB 1381 GAAGAAGCTCTCAATGAACAGTATGCGGGGTGTCTGCCACCCCTGAGAGGCGAGAGCAT 1440
QY 1441 GCCACCACTCCGTGGAGAGTCCAGAGAGGCCAGAGAGTCTGGGGGACCAATGGGTGCG 1500
DB 1441 GCCACCACTCCGTGGAGAGTCCAGAGAGGCCAGAGAGTCTGGGGGACCAATGGGTGCG 1500
QY 1501 CATGGAGGCGCGTGGAGAGATAGAGAGGCTTCCCTCCAAAGAGACCCCGGGGTTCCTCG 1560
DB 1501 CATGGAGGCGCGTGGAGAGATAGAGAGGCTTCCCTCCAAAGAGACCCCGGGGTTCCTCG 1560
QY 1561 AGGGAACCCCTCTGGAGAGAGAAAGCTCCAGCACCGAGCTGGAGACTGGGATGTCCCAA 1620
DB 1561 AGGGAACCCCTCTGGAGAGAGAAAGCTCCAGCACCGAGCTGGAGACTGGGATGTCCCAA 1620
QY 1621 TCGGGGTTTGGAACCAAGAACTTCCCTGAGAGACAGATGGCAACAGTGAAGGCCCC 1680
DB 1621 TCGGGGTTTGGAACCAAGAACTTCCCTGAGAGACAGATGGCAACAGTGAAGGCCCC 1680
QY 1681 AAAGCCTGAAGGCTTCTCCCGCAACCTTTCGCGCCCGGGGTGTGATGCTGGCAGAGG 1740
DB 1681 AAAGCCTGAAGGCTTCTCCCGCAACCTTTCGCGCCCGGGGTGTGATGCTGGCAGAGG 1740
QY 1741 TGGCCCTGTGGGCAATGCGGGAGAGAGAGGTGCTCATGATGCTGTGTGCTCCGGTGG 1800
DB 1741 TGGCCCTGTGGGCAATGCGGGAGAGAGAGGTGCTCATGATGCTGTGTGCTCCGGTGG 1800
QY 1801 AATGTTCAAGAGTGGCGGTGTGAGACAGAGGTGCTTCCGTGTGTGCGGGGCAATGGA 1860
DB 1801 AATGTTCAAGAGTGGCGGTGTGAGACAGAGGTGCTTCCGTGTGTGCGGGGCAATGGA 1860
QY 1861 CCGAGGTGCTTGTGTGAGAGAAAGAGAGTGGCTTGGGGGCCCCCTGAGACTTTGAT 1920
DB 1861 CCGAGGTGCTTGTGTGAGAGAAAGAGAGTGGCTTGGGGGCCCCCTGAGACTTTGAT 1920
QY 1921 GGAACAGATGGAGAGAAAGAGAGAGAGCTGGAGACCTGGAAAAATGGATAAAGCGA 1980
DB 1921 GGAACAGATGGAGAGAAAGAGAGAGAGCTGGAGACCTGGAAAAATGGATAAAGCGA 1980
QY 1981 GCAACCGTCAAGAGAGCGAGAGATCGGCGCTTACTAGATGCAAGACCCCGCAGACTGCA 2040
DB 1981 GCAACCGTCAAGAGAGCGAGAGATCGGCGCTTACTAGATGCAAGACCCCGCAGACTGCA 2040
QY 2041 GACTACCAAGATTTATTTTAAACAGAAATGTTTAAATTTATATTCATATTTATA 2100
DB 2041 GACTACCAAGATTTATTTTAAACAGAAATGTTTAAATTTATATTCATATTTATA 2100
QY 2101 ATGTTGGCCAACATTTATGATTTCTTGTCTGTAATTTAGATTTTCAACATTTGT 2160
DB 2101 ATGTTGGCCAACATTTATGATTTCTTGTCTGTAATTTAGATTTTCAACATTTGT 2160
QY 2161 GAAGAAACATTTAAACAGTTAAATGTTAGTGTGCGAGTTTCTTCTCTCTT 2220
DB 2161 GAAGAAACATTTAAACAGTTAAATGTTAGTGTGCGAGTTTCTTCTCTCTT 2220
QY 2221 AAAAATGTTTAAAGCTTTAAACATTTAAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2280
DB 2221 AAAAATGTTTAAAGCTTTAAACATTTAAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2280

```

```

QY 2281 GTGAGAACCAAGAGGCGCTTTAACTGTAACATGTCATGTTGATGTTTTTTT 2340
DB 2281 GTGAGAACCAAGAGGCGCTTTAACTGTAACATGTCATGTTGATGTTTTTTT 2340
QY 2341 TTTTAAATTAATTAATTTCAATGTTTAAATTAATTAATTAATTAATTAATTAAT 2390
DB 2341 TTTTAAATTAATTAATTTCAATGTTTAAATTAATTAATTAATTAATTAATTAAT 2390

RESULT 6
AAQ50643
ID AAQ50643 standard; cDNA; 2371 BP.
XX
AC AAQ50643;
XX
DT 25-MAR-2003 (revised)
DT 26-MAY-1994 (first entry)
XX
DE Human Ews gene clone BFLAC5 from foetal brain cDNA library.
XX
KW chromosomal translocation; chimeric; chimeric; Ewing sarcoma; Ews gene;
KW malignant melanoma; hum-fl-1;
KW primitive peripheral neuroectodermal tumour; human chromosome 11;
KW human chromosome 22; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 25..1995
FT FT /tag= a
FT FT /product= "EWS_Protein"
FT FT /transl_except= pos:1729..1731; aa:Val
FT polyA_signal 2143..2148
FT polyA_signal /tag= b
FT FT 2331..2336
FT FT /tag= c
XX
PN WO9323549-A2.
XX
PD 25-NOV-1993.
XX
PF 19-MAY-1993; 93WO-FR000494.
XX
PR 20-MAY-1992; 92FR-00006123.
XX
PA (CNRS ) CNRS CENT NAT RECH SCI.
XX
PI Aurias A, Delattre O, Desmaze C, Melot T, Peter M, Plougaestel B,
PI Thomas G, Zucman J;
XX
DR WPI; 1993-386580/48.
DR P-PSDB; AAR44555.
XX
PT New nucleic acid of EWS gene and its hybrid(s) - contg. gene sequence
PT involved in chromosomal trans-location, also derived mRNA, probes, fusion
PT proteins etc., for diagnosis and treatment of Ewing sarcoma and melanoma.
XX
PS Disclosure; Fig 6; 123pp; French.
XX
CC The probes 22R3 and 22R12 were used to screen a human foetal brain cDNA
CC library (Stratagene cat. # 936206). The clone BFLAC5 was identified and
CC sequenced. It represents the entire coding region and 3'-UTR of the Ews
CC gene. (Updated on 25-MAR-2003 to correct PN field.)
XX
SO Sequence 2371 BP; 639 A; 587 C; 659 G; 486 T; 0 U; 0 Other;

Query Match 99.1%; Score 2369.4; DB 2; Length 2371;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2370; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 20 GAGAACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 79
DB 1 GAGAACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 60

```

QY 80 GAGGCGACAGAGGCTACAGTGTCTTACACGCCACAGCCCACTCAAGATATGCAAGACC 139  
Db 61 GAGGCGACAGAGGCTACAGTGTCTTACACGCCACAGCCCACTCAAGATATGCAAGACC 120  
QY 140 ACCGAGGCTATATGGGCAAAAGCTATGGAACCTATGACAGGCCACTGATGTCACTAT 199  
Db 121 ACCGAGGCTATATGGGCAAAAGCTATGGAACCTATGACAGGCCACTGATGTCACTAT 180  
QY 200 ACCGAGGCTCAAGACCACTGCAACTATATGGGCAAGCCGCTATGCAACTTCTTATGGAAG 259  
Db 181 ACCGAGGCTCAAGACCACTGCAACTATATGGGCAAGCCGCTATGCAACTTCTTATGGAAG 240  
QY 260 CCTCCCACTGTTATCTACTCTCAACTGCCCCCAGGCAATACAGCCAGGCTGTCCAGGGG 319  
Db 241 CCTCCCACTGTTATCTACTCTCAACTGCCCCCAGGCAATACAGCCAGGCTGTCCAGGGG 300  
QY 320 TATGCACTGATGCTTATGATACCACTGCTACAGTCAACCAAGGCTCTCTAT 379  
Db 301 TATGCACTGATGCTTATGATACCACTGCTACAGTCAACCAAGGCTCTCTAT 360  
QY 380 GCACTCACTGCTATATGCACTCAAGCTGCTTATCCAGCTATATGGGCAAGCCAGCA 439  
Db 361 GCACTCACTGCTATATGCACTCAAGCTGCTTATCCAGCTATATGGGCAAGCCAGCA 420  
QY 440 GCCACTGCACTCAAGACCCGAGATGGAACAAGCCCACTGAGACTAGTCAACTCA 499  
Db 421 GCCACTGCACTCAAGACCCGAGATGGAACAAGCCCACTGAGACTAGTCAACTCA 480  
QY 500 TCTAGCAAGGGGTTTCAACACAGCCAGCTATGATGATGACAGATTACTACAGTAT 559  
Db 481 TCTAGCAAGGGGTTTCAACACAGCCAGCTATGATGATGACAGATTACTACAGTAT 540  
QY 560 CCCCAAGTACTTGGGAGCTACCCCACTGCAAGCTGCACTTCACTTCACTTCACTT 619  
Db 541 CCCCAAGTACTTGGGAGCTACCCCACTGCAAGCTGCACTTCACTTCACTTCACTT 600  
QY 620 ACCAGTATCTCTCTCAACAGCCGAGCTAGTATGATCAGAGGAGTAACTCTCAGAGAA 679  
Db 601 ACCAGTATCTCTCTCAACAGCCGAGCTAGTATGATGATGAGAGTAACTCTCAGAGAA 660  
QY 680 ACCATATGGGCAACCGAGAGCTATGACAGCAGAGTATGATGATGATGATGATGAT 739  
Db 661 ACCATATGGGCAACCGAGAGCTATGACAGCAGAGTATGATGATGATGATGATGAT 720  
QY 740 GGGCAGAGGCTCTCCACTAGTATCCCACTGCAAGCTGCAAGCTCTCAAGT 799  
Db 721 GGGCAGAGGCTCTCCACTAGTATCCCACTGCAAGCTGCAAGCTCTCAAGT 780  
QY 800 CAATATAGCCCAAGAGAGCAGCTACGAGCAGAGGTTCACTCCGACAGAGCAACCC 859  
Db 781 CAATATAGCCCAAGAGAGCAGCTACGAGCAGAGGTTCACTCCGACAGAGCAACCC 840  
QY 860 AGTAGATGGGTTTATATGGGCAAGTCTGAGGATTTTCCGAGCCAGAGAGAAACCG 919  
Db 841 AGTAGATGGGTTTATATGGGCAAGTCTGAGGATTTTCCGAGCCAGAGAGAAACCG 900  
QY 920 AGCATGATGGCCCTGATTAACCGGGGCAAGGGGGAATTTGATCTGTGAGGAGT 979  
Db 901 AGCATGATGGCCCTGATTAACCGGGGCAAGGGGGAATTTGATCTGTGAGGAGT 960  
QY 980 AGCAGAGTGGGGGGGGAGAGACCGGCTGGAATGGGCAAGGCTGAGAGAGGAGT 1039  
Db 961 AGCAGAGTGGGGGGGGAGAGACCGGCTGGAATGGGCAAGGCTGAGAGAGGAGT 1020  
QY 1040 TTCAATAGGCTGTGAGACCATGATGAAAGACAGATCTTATGAGCCCACTGTA 1099  
Db 1021 TTCAATAGGCTGTGAGACCATGATGAAAGACAGATCTTATGAGCCCACTGTA 1080  
QY 1100 GATTCAGATGAAGCTTGAACAAGTGAATTTATATGCAAGGATTTAAATGACAGTGT 1159  
Db 1081 GATTCAGATGAAGCTTGAACAAGTGAATTTATGTAAGGATTTAAATGACAGTGT 1140

QY 1160 ACTCTAGATGATCTGGCAGACTTCTTTAAGCAGTGTGGGTTGTTAAGATGAACAAGAGA 1219  
Db 1141 ACTCTAGATGATCTGGCAGACTTCTTTAAGCAGTGTGGGTTGTTAAGATGAACAAGAGA 1200  
QY 1220 ACTGGGCAACCCATGATCCATCTTACCTGACAGAGAAACAGAAAGCCCAAGCCAT 1279  
Db 1201 ACTGGGCAACCCATGATCCATCTTACCTGACAGAGAAACAGAAAGCCCAAGCCAT 1260  
QY 1280 GGCAGAGTCTCTAAGAGACCCCACTGACAGAGGCTGCGGGAATGTTGATGGG 1339  
Db 1261 GGCAGAGTCTCTAAGAGACCCCACTGACAGAGGCTGCGGGAATGTTGATGGG 1320  
QY 1340 AAAAGATTTCAAGGAGCAACTTAAAGTCTCTGCTGGAAGAGCCCTCAATGAAC 1399  
Db 1321 AAAAGATTTCAAGGAGCAACTTAAAGTCTCTGCTGGAAGAGCCCTCAATGAAC 1380  
QY 1400 AGTATGCGGGGTGTCTGCACTCCGCTGAGGCAAGGATGCCACCACTCCGTGA 1459  
Db 1381 AGTATGCGGGGTGTCTGCACTCCGCTGAGGCAAGGATGCCACCACTCCGTGA 1440  
QY 1460 GGTCCAGAGAGGCGCCAGAGGCTCTGGGGGACCAATGGGTCGATGGGAGGCGTGAAGA 1519  
Db 1441 GGTCCAGAGAGGCGCCAGAGGCTCTGGGGGACCAATGGGTCGATGGGAGGCGTGAAGA 1500  
QY 1520 GATAGAGAGGCTTCTCTCAAGAGACCCGAGGTTCCCGAGGAAACCTCTGAGAGA 1579  
Db 1501 GATAGAGAGGCTTCTCTCAAGAGACCCGAGGTTCCCGAGGAAACCTCTGAGAGA 1560  
QY 1580 GGAAGAGTCTGACACCCAGCTGGAAGCTGGCACTGTCCCAATCCGGGTTGTGAACCAAG 1639  
Db 1561 GGAAGAGTCTGACACCCAGCTGGAAGCTGGCACTGTCCCAATCCGGGTTGTGAACCAAG 1620  
QY 1640 AACTTGAGTGAAGAACAGAGTGAACAGAGTGAAGGAGGAGGAGGAGGAGGAGGAGGAG 1699  
Db 1621 AACTTGAGTGAAGAACAGAGTGAACAGAGTGAAGGAGGAGGAGGAGGAGGAGGAGGAG 1680  
QY 1700 CCGCACCCTTCCGCGCCCGGGGAGTGTATCTGTGAGAGAGTGTGAGTGTGAGTGTGAG 1759  
Db 1681 CCGCACCCTTCCGCGCCCGGGGAGTGTATCTGTGAGAGAGTGTGAGTGTGAGTGTGAG 1740  
QY 1760 GAGAGAGAGTGTGAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1819  
Db 1741 GAGAGAGAGTGTGAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1800  
QY 1820 GGTGAGAGCAGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGT 1879  
Db 1801 GGTGAGAGCAGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGT 1860  
QY 1880 GGAAGAGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 1939  
Db 1861 GGAAGAGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 1920  
QY 1940 AGAGAGAGCAGTGAAGACCTGGAAGAAATGATTAAGGCGAGCAGCTCAAGAGCCGAGA 1999  
Db 1921 AGAGAGAGCAGTGAAGACCTGGAAGAAATGATTAAGGCGAGCAGCTCAAGAGCCGAGA 1980  
QY 2000 GATCGGCCCTAATAAGTGAAGACCCGAGAGGCTGATGATGATGATGATGATGATGAT 2059  
Db 1981 GATCGGCCCTAATAAGTGAAGACCCGAGAGGCTGATGATGATGATGATGATGATGAT 2040  
QY 2060 TAAACAGAGAAATGTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2119  
Db 2041 TAAACAGAGAAATGTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2100  
QY 2120 GATTAATCTTGTCTGATCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2179  
Db 2101 GATTAATCTTGTCTGATCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2160  
QY 2180 TTAATGATAGTGTGAGGAGTGTGATGATGATGATGATGATGATGATGATGATGATGAT 2239  
Db 2161 TTAATGATAGTGTGAGGAGTGTGATGATGATGATGATGATGATGATGATGATGATGAT 2220  
QY 2240 TTTAACAATGGGAACCTTGTGAGATGCTCAGATATGATGAGAAACCAAGAGGCC 2299

	2221	2300	2281	2360	2341
Db	TTTAAACATGGGAAACCCCTTGTGACGATCTCAGATCATGTGGAGAACCAAGGGCC				
Qy		TCTTAACTGTAACAAATGTCATGCTGTGATGTTTTTTTTTTTTTAAATAAAATTC			
Db		2281	TCTTAACTGTAACAAATGTCATGCTGTGATGTTTTTTTTTTTTTAAATAAAATTC		
Qy				2360	
Db				2341	

## RESULT 7

AD086032 standard, cDNA; 2372 BP.  
AD086032,  
07-OCT-2004 (first entry)  
Human tumour-associated antigenic target (TAT) cDNA sequence #2904.  
human, tumour-associated antigenic target; TAT; cytostatic; gene therapy;  
cancer; cell proliferative disorder; gene; ss.  
Homo sapiens.  
W02004060270-A2.  
22-JUL-2004.  
15-OCT-2003; 2003WO-US029126.  
18-OCT-2002; 2002US-0418988P.  
(GENTH) GENENTECH INC.  
(MOTD/) WU T D.  
(ZHOU/) ZHOU Y.  
Wu TD, Zhou Y;  
WPI; 2004-534300/51.  
New nucleic acid molecule and encoded polypeptide, for diagnosing,  
preventing or treating cell proliferative disorders such as cancer.  
Claim 1; SEQ ID NO 2904; 5504bp; English.



Db	781	CTTAAAGCCAGACTCCAGTCAATTATAGCCAAACAGACGACGCTTACGGGACGACGATTC	840
Qy	841	ATTCCGACGAGACCAACCCCACTAGCATGGAATGTTATGCGAGGAGTCTGAGGATTTTC	900
Db	841	ATTCCGACGAGACCAACCCCACTAGCACTGGGTTTATGCGAGGAGTCTGAGGATTTTC	900
Qy	901	CGGACCAAGGAGAAACCGGACATGAGTGGCCCTGATPACCGGGGACAGGGGAGAGGGGG	960
Db	901	CGGACCAAGGAGAAACCGGACATGAGTGGCCCTGATPACCGGGGACAGGGGAGAGGGGG	960
Qy	961	ATTGATGATGATGAGGACATGACAGAGGAGGCGGGGAGAGAGAGCGCGGTGAATGGGACG	1020
Db	961	ATTGATGATGATGAGGACATGACAGAGGAGGCGGGGAGAGAGAGCGCGGTGAATGGGACG	1020
Qy	1021	CGCTGAGAGCGAGGTGGCTTCAATAAGCCTGTGTGACCCATGTGATGAGGACCAAGTCT	1080
Db	1021	CGCTGAGAGCGAGGTGGCTTCAATAAGCCTGTGTGACCCATGTGATGAGGACCAAGTCT	1080
Qy	1081	TGATCTAGAGCCCACTGTGATPCCAGATGANGACTGTGACCAAGTGCATTTATGTGACA	1140
Db	1081	TGATCTAGAGCCCACTGTGATPCCAGATGANGACTGTGACCAAGTGCATTTATGTGACA	1140
Qy	1141	AGGATTAATGACAGTGTGACTGTAGATGATCTGGACGACTTCTTAAAGCAGTGTGGAGT	1200
Db	1141	AGGATTAATGACAGTGTGACTGTAGATGATCTGGACGACTTCTTAAAGCAGTGTGGAGT	1200
Qy	1201	TGTTAAGATGAAACAAGAGAACTGGGCAACCCATGATCCAATCTTACCTGACAAAGAAAC	1260
Db	1201	TGTTAAGATGAAACAAGAGAACTGGGCAACCCATGATCCAATCTTACCTGACAAAGAAAC	1260
Qy	1261	AGGAAAGCCCAAGGCGATGCCACAAGTGTCTTATGAGAACCCACCTGCCAAGGCTGC	1320
Db	1261	AGGAAAGCCCAAGGCGATGTGCACAGTGTCTTATGAGAACCCACCTGCCAAGGCTGC	1320
Qy	1321	CGTGAATGATGTTGATGGGAAAGATTTTCAAGGGAGGCAAACTTAAAGTCTCCCTTGCCTG	1380
Db	1321	CGTGAATGATGTTGATGGGAAAGATTTTCAAGGGAGGCAAACTTAAAGTCTCCCTTGCCTG	1380
Qy	1381	GAAAGAGCCTCCAAATGAAACAGTATGCGGGGTGTCTGCCACCCCGTGAAGGCGACAGGCAT	1440
Db	1381	GAAAGAGCCTCCAAATGAAACAGTATGCGGGGTGTCTGCCACCCCGTGAAGGCGACAGGCAT	1440
Qy	1441	GCCACCAACAATCCGTGGAAGGTCCAAGAGGCCCAAGAGGTCTTGGGGGACCCATGGGTGC	1500
Db	1441	GCCACCAACAATCCGTGGAAGGTCCAAGAGGCCCAAGAGGTCTTGGGGGACCCATGGGTGC	1500
Qy	1501	CATGGGAGGCGGTGAGGAGATPAGAGGAGCTTCCCTCAAGAGAGACCCCGGGGTTCCTCG	1560
Db	1501	CATGGGAGGCGGTGAGGAGATPAGAGGAGCTTCCCTCAAGAGAGACCCCGGGGTTCCTCG	1560
Qy	1561	AGGGAACCCCTCTGAGAGGAGAAACGTCCAGACCCAGAGCTGGAAGCTGGACTGTCCCAA	1620
Db	1561	AGGGAACCCCTCTGAGAGGAGAAACGTCCAGACCCAGAGCTGGAAGCTGGACTGTCCCAA	1620
Qy	1621	TCCGGGTTGTGAAAACAGAACTTCCGCTGAGAAACAGAGTGCACCAAGTGAAGGCCCC	1680
Db	1621	TCCGGGTTGTGAAAACAGAACTTCCGCTGAGAAACAGAGTGCACCAAGTGAAGGCCCC	1680
Qy	1681	AAAGCTTGAAGGCTTCTCCCGCCACCTTTCCGCCCCCGGGGTGTATCTGGACAAGG	1740
Db	1681	AAAGCTTGAAGGCTTCTCCCGCCACCTTTCCGCCCCCGGGGTGTATCTGGACAAGG	1740
Qy	1741	TGGCCCTGTGAGCATGCGGGAGAGAAAGTGTGCTCATGTGATGTGTGCTCCCGGTGG	1800
Db	1741	TGGCCCTGTGAGCATGCGGGAGAGAAAGTGTGCTCATGTGATGTGTGCTCCCGGTGG	1800
Qy	1801	AATGTTCAAGAGTGGCCGTGTGTGAGACAGAGTGTGCTTCCGTGTGTGCGGGGACATGGA	1860
Db	1801	AATGTTCAAGAGTGGCCGTGTGTGAGACAGAGTGTGCTTCCGTGTGTGCGGGGACATGGA	1860
Qy	1861	CGAGAGTGGCTTTGTGTGAGAGAGACAGAGTGTGCTTGGGGGGCCCTGTGACCTTTGAT	1920
Db	1861	CGAGAGTGGCTTTGTGTGAGAGAGACAGAGTGTGCTTGGGGGGCCCTGTGACCTTTGAT	1920

Db	1661	TCGAGGTGTCCTTGCTGTGAGAAAGACGAGGTGCCTCTGGGGGGCCCTCTGGACCTTGTAT	1920
QY	1291	GGAAACAATGTGGAGAAAGAAAGAGAGACGCTGGAGACCTCGAAAAATGGATTAAGGCCGA	1980
Db	1921	GGAAACAATGTGGAGAAAGAAAGAGAGACGCTGGAGACCTCGAAAAATGGATTAAGGCCGA	1980
QY	1981	GCACCGTCAGAGCGCAGAGAGTCGCCCTCTACTAGATCAGAGACCCCGCAGAGCTGCATT	2040
Db	1981	GCACCGTCAGAGCGCAGAGAGTCGCCCTCTACTAGATCAGAGACCCCGCAGAGCTGCATT	2040
QY	2041	GACTAACCAAGATTATTTTTTAAACAGAAATGTTTTAAATTATTAATTCGATATTTATA	2100
Db	2041	GACTAACCAAGATTATTTTTTAAACAGAAATGTTTTAAATTATTAATTCGATATTTATA	2100
QY	2101	ATGTGGCCACAATTATGATTTTCCCTGTCGTACTTAGATTTTTTACCATTGTG	2160
Db	2101	ATGTGGCCACAATTATGATTTTCCCTGTCGTACTTAGATTTTTTACCATTGTG	2160
QY	2161	GAAGAAACATTAAAAACAAGTTAAATGATGTGTGCGAGTTTTTTTTTCTTCTCTTT	2220
Db	2161	GAAGAAACATTAAAAACAAGTTAAATGATGTGTGCGAGTTTTTTTTTCTTCTCTTT	2220
QY	2221	AAAAATGCTGTTTAAAGCTTTAACANGGGAAACCCCTGGAGACATGCTCAGTACATT	2280
Db	2221	AAAAATGCTGTTTAAAGCTTTAACANGGGAAACCCCTGGAGACATGCTCAGTACATT	2280
QY	2281	GTGGAGAACCAAGAGGCGCTCTTAAGTAACTGTAACTGTCAATGTTGATGTGATGTTTTTTTT	2340
Db	2281	GTGGAGAACCAAGAGGCGCTCTTAAGTAACTGTAACTGTCAATGTTGATGTGATGTTTTTTTT	2340
QY	2341	TTTTTTTAAATTAATTAATTCCAAATGTTAATA 2373	
Db	2341	TTTTTTT-AAATTAATTAATTCCAAATGTTAATA 2372	

XX	RESULT 8
AA570647	
ID	AA570647 standard; cDNA, 2388 BP.
XX	
XX	AA570647,
XX	
DT	13-FEB-2002 (first entry)
XX	
DE	DNA encoding novel human diagnostic protein #6451.
KM	Human; chromosome mapping; gene mapping; forensic;
KW	food supplement; medical imaging; diagnostic; genetic disorder; ss
XX	
OS	Homo sapiens.
PN	WO200175067-A2.
PD	
XX	11-OCT-2001.
XX	
PF	30-MAR-2001; 2001WO-US008631.
XX	
PR	31-MAR-2000; 2000US-00540217.
PR	23-AUG-2000; 2000US-00645167.
XX	
PA	(HYSE-) HYSEQ INC.
XX	
PI	Dmanac RT, Liu C, Tang YT;
XX	
DR	WPI; 2001-639362/73.
DR	P-PSDB; ABG06460.
XX	
XX	
PT	New isolated polynucleotide and encoded polypeptides, useful in
PT	diagnostics, forensics, gene mapping, identification of mutations
PT	responsible for genetic disorders or other traits and to assess
PT	biodiversity.
XX	
PS	Claim 1; SEQ ID NO 6451; 103bp; English.
XX	







QY 702 ATGACAGCAGAGTACTATGTCACAAACAGCTATGGGACAGACCTCCACTAGTT 761  
 DB 1608 ATGACAGCAGAGTACTATGTCACAAACAGCTATGGGACAGACCTCCACTAGTT 1549  
 QY 762 ACCACCCCAAACTGATCTCTACAGCCAGCTCCAACTCAATATAGCCAAACAGAGCAGA 821  
 DB 1548 ACCACCCCAAACTGATCTCTACAGCCAGCTCCAACTCAATATAGCCAAACAGAGCAGA 1489  
 QY 822 GCTACGGGACAGAGTTCAATCCGACAGGACCAACCCAGTAGCATGGGTGTTATGGC 881  
 DB 1488 GCTACGGGACAGAGTTCAATCCGACAGGACCAACCCAGTAGCATGGGTGTTATGGC 1429  
 QY 882 AGGAGCTGAGAGATTTTCCGACAGGACCAACCCAGTAGCATGGGTGTTATGGC 941  
 DB 1428 AGGAGCTGAGAGATTTTCCGACAGGACCAACCCAGTAGCATGGGTGTTATGGC 1369  
 QY 942 GGGGACAGGAGAGAGGGGAGTTGATGCTGAGGACATGACAGAGGTGGGCGGAGAGAG 1001  
 DB 1368 GGGGACAGGAGAGAGGGGAGTTGATGCTGAGGACATGACAGAGGTGGGCGGAGAGAG 1309  
 QY 1002 GACGCGGTGAGATGGGACGCTGAGAGAGGAGGCTTCAATAGCTGTGAGACCA 1061  
 DB 1308 GACGCGGTGAGATGGGACGCTGAGAGAGGAGGCTTCAATAGCTGTGAGACCA 1252  
 QY 1062 TGGATGAAGACCAAGATCTGATCTAGGACCACTGATGATCCAGATGAAGACTGTGACA 1121  
 DB 1251 TGGATGAAGACCAAGATCTGATCTAGGACCACTGATGATCCAGATGAAGACTGTGACA 1192  
 QY 1122 ACAGTCAATTTTATGATCAAGATTTAAATGACAGTGAATCTAGATGATCTGACAGACT 1181  
 DB 1191 ACAGTCAATTTTATGATCAAGATTTAAATGACAGTGAATCTAGATGATCTGACAGACT 1132  
 QY 1182 TCTTTAAGCAGTGTGGGTTGTTAAGATGAACAAAGAACTGGGCAACCCATGATCCACA 1241  
 DB 1131 TCTTTAAGCAGTGTGGGTTGTTAAGATGAACAAAGAACTGGGCAACCCATGATCCACA 1072  
 QY 1242 TCTAATCTGACAAAGAAACAGAAAGCCCAAGGCGCATGATGCTTATGAAGACC 1301  
 DB 1071 TCTAATCTGACAAAGAAACAGAAAGCCCAAGGCGCATGATGCTTATGAAGACC 1012  
 QY 1302 CACCCACTGCGCAAGGCTGCGGTGGAATGTTTGAAGGGAAGATTTTCAAGGAGCAAC 1361  
 DB 1011 CACCCACTGCGCAAGGCTGCGGTGGAATGTTTGAAGGGAAGATTTTCAAGGAGCAAC 952  
 QY 1362 TTAAGTCTCTCTTGTCTCGAAGAAAGCTTCAATGAACAGTATGCGGGTGTCTGACAC 1421  
 DB 951 TTAAGTCTCTCTTGTCTCGAAGAAAGCTTCAATGAACAGTATGCGGGTGTCTGACAC 892  
 QY 1422 CCGGTGAGGACAGAGCATGCTCCACACACTCCGTGAGGATCCAGAGGCGCCAGAGGTC 1481  
 DB 891 CCGGTGAGGACAGAGCATGCTCCACACACTCCGTGAGGATCCAGAGGCGCCAGAGGTC 832  
 QY 1482 CTGGGAGAACCATGGGTGCGCATGAGAGGCGGTGAGAGAGATAGAGAGGCTTCCCTCAA 1541  
 DB 831 CTGGGAGAACCATGGGTGCGCATGAGAGGCGGTGAGAGAGATAGAGAGGCTTCCCTCAA 772  
 QY 1542 GAGGACCCCGGGGTTCCCGAGGAAACCCCTCTGAGAGGAGAAAGTCCAGACCGAGCTG 1601  
 DB 771 GAGGACCCCGGGGTTCCCGAGGAAACCCCTCTGAGAGGAGAAAGTCCAGACCGAGCTG 712  
 QY 1602 GAGACTGGCAGTGTCCCAATCCGGGTTTGGAAAACCAAACTTCCGCTGAGAAACAGAGT 1661  
 DB 711 GAGACTGGCAGTGTCCCAATCCGGGTTTGGAAAACCAAACTTCCGCTGAGAAACAGAGT 652  
 QY 1662 GCAACCAAGTGAAGGCCCAAAAGCTGAAGGCTTCCCGGCAACCTTCCGCGCCCGG 1721  
 DB 651 GCAACCAAGT----- 643  
 QY 1722 GTGGTATCGTGGCAGAGGTGCGCTGTGTGCACTGGGAGAGAAAGTGGCTCTAGT 1781  
 DB 642 GTGGTATCGTGGCAGAGGTGCGCTGTGTGCACTGGGAGAGAAAGTGGCTCTAGT 583

QY 1782 ATGCTGTGTGCTCCCGGTGAATGTTCAAGAGTGGCCGTGTGAGACAGAGGTGCTTCC 1841  
 DB 582 ATGCTGTGTGCTCCCGGTGAATGTTCAAGAGTGGCCGTGTGAGACAGAGGTGCTTCC 523  
 QY 1842 GTGGTGGCCGGGACATGACCCGAGTGTGCTTGTGTGAGAGAAACAGAGTGGCTTGGAG 1901  
 DB 522 GTGGTGGCCGGGACATGACCCGAGTGTGCTTGTGTGAGAGAAACAGAGTGGCTTGGAG 463  
 QY 1902 GGGCCCTGTGACCTTGTATGAAACAGATGAGAGAAAGAGAGAGAGCTGTGAGACCTG 1961  
 DB 462 GGGCCCTGTGACCTTGTATGAAACAGATGAGAGAAAGAGAGAGAGCTGTGAGACCTG 403  
 QY 1962 GAAAAATGATTAAGGCGACACCGTCAAGAGGCGAGATGCGCCCTACTATATGACA 2021  
 DB 402 GAAAAATGATTAAGGCGACACCGTCAAGAGGCGAGATGCGCCCTACTATATGACA 343  
 QY 2022 GACCCCGCAGAGCTGATGATGATCAACAGATTTATTTTAAACAGAAATGTTTAAAT 2081  
 DB 342 GACCCCGCAGAGCTGATGATGATCAACAGATTTATTTTAAACAGAAATGTTTAAAT 283  
 QY 2082 TTAATATTCATATTTAATATGTTGCGCAACAACTATATGATTTGCTGTGATCTT 2141  
 DB 282 TTAATATTCATATTTAATATGTTGCGCAACAACTATATGATTTGCTGTGATCTT 223  
 QY 2142 AGTATTTTCAACATTTGTGAGAAACATTTAAACAAAGTTAAATGTAAGTGGAGATT 2201  
 DB 222 AGTATTTTCAACATTTGTGAGAAACATTTAAACAAAGTTAAATGTAAGTGGAGATT 163  
 QY 2202 TTTTTCCTCTCTCTTTTAAATAGTTGTTAAGACTTTAACAATGGGAAACCCCTGT 2261  
 DB 162 TTTTTCCTCTCTCTTTTAAATAGTTGTTAAGACTTTAACAATGGGAAACCCCTGT 103  
 QY 2262 GAGCATGCTCAATATCATTTGTGAGAAACAGAGGCTCTTAACTGTAACTATGTTTAT 2321  
 DB 102 GAGCATGCTCAATATCATTTGTGAGAAACAGAGGCTCTTAACTGTAACTATGTTTAT 43  
 QY 2322 GGTGTGATGTTTGT 2363  
 DB 42 GGTGTGATGTTTGT 1

RESULT 10  
 AAS62262/C  
 ID AAS62262 standard; cDNA; 2176 BP.  
 XX  
 AC AAS62262;  
 XX  
 DT 14-FEB-2002 (first entry)  
 XX  
 DE cDNA sequence #49 encoding novel human secreted protein.  
 XX  
 KW Human secreted protein; hyperproliferative disorder; autoimmune disorder;  
 KW immune deficiency disorder; blood disorder; inflammatory disorder;  
 KW infectious disorder; gene therapy; antimicrobial; hepatotropic;  
 KW immunosuppressive; antineumatic; ss.  
 OS  
 OS Homo sapiens.  
 PN MO200177291-A2.  
 XX  
 PD 18-OCT-2001.  
 XX  
 PD 29-MAR-2001; 2001WO-US010485.  
 PF  
 XX 06-APR-2000; 2000US-0195604P.  
 PR  
 XX (GENY ) GENETICS INST INC.  
 PA  
 XX Wong GG, Clark HF, Fechtel K, Agostino MJ, Howes SH, Reenick RU;  
 PI Gulukota K, Graham JR;  
 XX  
 DR MPI; 2002-010900/01.  
 XX

PT New polynucleotides encoding secreted proteins useful for treating e.g.  
 asthma, HIV and Crohn's disease.

Claim 1; Page 100; 391pp; English.

CC The present invention relates to the isolation of novel cDNA sequences  
 CC which encode human secreted proteins. The cDNA sequences have been  
 CC derived from a variety of human tissues. The invention also provides a  
 CC method for producing proteins from these polynucleotide sequences. The  
 CC proteins are useful for identifying compounds that modulate their  
 CC activity and production, and the cell is also useful for identifying  
 CC compounds that modulate expression of the polynucleotide sequences  
 CC encoding the secreted proteins. The sequences of the invention are useful  
 CC for treating diseases such as hyperproliferative disorders (e.g. cancer),  
 CC immune deficiency disorders (e.g. severe combined immunodeficiency  
 CC (SCID)), autoimmune disorders (e.g. multiple sclerosis), blood disorders  
 CC (e.g. thrombocytopaenia), inflammatory disorders (e.g. arthritis) and  
 CC infectious disorders (e.g. hepatitis). The polynucleotide sequences of  
 CC the invention are also useful in gene therapy. AAS62214-AAS62838  
 CC represent the cDNA sequences of the invention that encode for novel human  
 CC secreted proteins

XX Sequence 2176 BP; 413 A; 617 C; 563 G; 583 T; 0 U; 0 Other;

Query Match 89.6%; Score 2141.8; DB 6; Length 2176;  
 Best Local Similarity 99.9%; Pred. No. 0;  
 Matches 2154; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

QY 33 AAGAGAGAAAATGCGCTCCACGAGATTACAGTACCTATAGCCAAAGCTGCAGCGCAGCAG 92  
 Db 2171 AAGGTAGAAAATGCGCTCCACGAGATTACAGTACCTATAGCCAAAGCTGCAGCGCAGCAG 2112  
 QY 93 GCTACAGTGGTTACACCGCCCAACCTCAAGATATGACAGACACCCAGGCTATAG 152  
 Db 2111 GCTACAGTGGTTACACCGCCCAACCTCAAGATATGACAGACACCCAGGCTATAG 2052  
 QY 153 GGGCAAAAAGCTATGGAACCTATGACAGCGCCACTAGTACCTATACCCAGGCTCAGA 212  
 Db 2051 GGGCAAAAAGCTATGGAACCTATGACAGCGCCACTAGTACCTATACCCAGGCTCAGA 1992  
 QY 213 CCACTGCACCTATGCGCAGACCGCCTATAGCACTTTATGACAGCGCTCCAGCTGGTT 272  
 Db 1991 CCACTGCACCTATGCGCAGACCGCCTATAGCACTTTATGACAGCGCTCCAGCTGGTT 1932  
 QY 273 ATATCTACTCCAACTGCCCCCAGCATACAGCGCTGTCCAGGGGTATGGAATGCTGCTG 332  
 Db 1931 ATATCTACTCCAACTGCCCCCAGCATACAGCGCTGTCCAGGGGTATGGAATGCTGCTG 1872  
 QY 333 CTATGATATACCACTGCTACAGTCAACCAACCGAGGCTCTCTATGACAGCTCAGTCTG 392  
 Db 1871 CTATGATATACCACTGCTACAGTCAACCAACCGAGGCTCTCTATGACAGCTCAGTCTG 1812  
 QY 393 CATATGCACTGACCTGCTTATCCAGCCTATGCGCAGCAGCAGCAGCAGCAGCAGCAG 452  
 Db 1811 CATATGCACTGACCTGCTTATCCAGCCTATGCGCAGCAGCAGCAGCAGCAGCAGCAG 1752  
 QY 453 CAAGACCGCAGATGGAAGAAACAGCCCACTGAGACTAGTCAACCTCAATCTAGACAGGGG 512  
 Db 1751 CAAGACCGCA-GATGGAAGAAACAGCCCACTGAGACTAGTCAACCTCAATCTAGACAGGGG 1693  
 QY 513 GTTACAAACAGCCAGCCTAGATATGAGACAGAGTAACTACAGTTATCCCAAGTATCTCT 572  
 Db 1692 GTTACAAACAGCCAGCCTAGATATGAGACAGAGTAACTACAGTTATCCCAAGTATCTCT 1633  
 QY 573 GAGAGTACCCCATGACAGCAGTCACTGACCTTCATCTCACTCTCTCAAGAGTATCTCT 632  
 Db 1632 GAGAGTACCCCATGACAGCAGTCACTTCATCTCACTCTCTCAAGAGTATCTCT 1573  
 QY 633 CTACACAGCCGACTAGTATGATCAAGAGCTTACTCTCAGCAGAAACCTATAGGGCAAC 692  
 Db 1572 CTACACAGCCGACTAGTATGATCAAGAGCTTACTCTCAGCAGAAACCTATAGGGCAAC 1513  
 QY 693 CGAGCAGCTATGACAGCAGAGTAGTATGATGCAAAAGCAGCTATGCGCAGAGCCTC 752

Db 1512 CGAGCAGCTATGACAGCAGAGTAGCTATGTCACAAAGCAGCTATGCGCAGCAGCCTC 1453  
 QY 753 CCACTAGTATCCCAACCCCAAACTGATCTCTACAGCCAGCTCCAACTCAATATAGCAAC 812  
 Db 1452 CCACTAGTATCCCAACCCCAAACTGATCTCTACAGCCAGCTCCAACTCAATATAGCAAC 1393  
 QY 813 AAGAGCAGCTACCGGGGAGCAGAGTTCCTCCGACAGGACCAACCCCAAGTATGAGTGGC 872  
 Db 1392 AAGAGCAGCTACCGGGGAGCAGAGTTCCTCCGACAGGACCAACCCCAAGTATGAGTGGC 1333  
 QY 873 TTTATGCGCAGAGTCTGAGGATTTTCGACACAGAGAGAACCCGAGCATGAGTGGC 932  
 Db 1332 TTTATGCGCAGAGTCTGAGGATTTTCGACACAGAGAGAACCCGAGCATGAGTGGC 1273  
 QY 933 CTGATPACCGGGGAGCAGGAGGAGGAGGAGTTCGTCGAGGACATGACAGAGTGGC 992  
 Db 1272 CTGATPACCGGGGAGCAGGAGGAGGAGGAGTTCGTCGAGGACATGAGAGTGGC 1213  
 QY 993 GGGGAGGAGGAGCAGCGGTGGAATGGGACGCTGAGAGCAGAGTGGCTCAATPAGCTG 1052  
 Db 1212 GGGGAGGAGGAGCAGCGGTGGAATGGGACGCTGAGAGCAGAGTGGCTCAATPAGCTG 1153  
 QY 1053 GTGACCCATGATGAGAGCAGATCTTGATCTAGGCCCACTGTAGATCCAGATGAAG 1112  
 Db 1152 GTGACCCATGATGAGAGCAGATCTTGATCTAGGCCCACTGTAGATCCAGATGAAG 1093  
 QY 1113 ACTTGACAAAGTGCATTTATGTCMAAGATTAATGACAGTGTGATCTTAGATGATC 1172  
 Db 1092 ACTTGACAAAGTGCATTTATGTCMAAGATTAATGACAGTGTGATCTTAGATGATC 1033  
 QY 1173 TGGCAGACTCTTTTAAGCAGTGTGGGGTGTTAAGATGAACAAGAACTGGGCAACCA 1232  
 Db 1032 TGGCAGACTCTTTTAAGCAGTGTGGGGTGTTAAGATGAACAAGAACTGGGCAACCA 973  
 QY 1233 TGATCCACTCTACCTGACCAAGAAACAGAAAGCCCAAGCGATGTCACAGTGTCT 1292  
 Db 972 TGATCCACTCTACCTGACCAAGAAACAGAAAGCCCAAGCGATGTCACAGTGTCT 913  
 QY 1293 ATGAAGACCCACCTGCTGCAAGAGCTGCGCTGTGAAATGTTTGAATGGGAAGATTTTCAAG 1352  
 Db 912 ATGAAGACCCACCTGCTGCAAGAGCTGCGCTGTGAAATGTTTGAATGGGAAGATTTTCAAG 853  
 QY 1353 GAGGCAAACTTAAAGTCTCCTTGTGCGAAGAGCTCCCAATGAAACAGTATCGGGGTG 1412  
 Db 852 GAGGCAAACTTAAAGTCTCCTTGTGCGAAGAGCTCCCAATGAAACAGTATCGGGGTG 793  
 QY 1413 GTCTGCACCCCGTGAAGGAGCAGAGCATGCCACCACTCCGTGAGGTCCAGAGGCC 1472  
 Db 792 GTCTGCACCCCGTGAAGGAGCAGAGCATGCCACCACTCCGTGAGGTCCAGAGGCC 733  
 QY 1473 CAGAGGCTCTGCGGAGACCAATGCGTGCATGAGAGGCGCTGAGAGATAGAGAGGCT 1532  
 Db 732 CAGAGGCTCTGCGGAGACCAATGCGTGCATGAGAGGCGCTGAGAGATAGAGAGGCT 673  
 QY 1533 TGCCTCCAGAGAGAACCCCGGGGTTCGAGGGGAAACCTCTGAGAGGAGGAAAGTCCAGC 1592  
 Db 672 TGCCTCCAGAGAGAACCCCGGGGTTCGAGGGGAAACCTCTGAGAGGAGGAAAGTCCAGC 613  
 QY 1593 ACCGAGCTGAGACTGCGAGTCCCAATCCGGGTGTGAGAAACAGAACTTGCTGTGA 1652  
 Db 612 ACCGAGCTGAGACTGCGAGTCCCAATCCGGGTGTGAGAAACAGAACTTGCTGTGA 553  
 QY 1653 GAAACAGTGCACCACTGTTAAGGCCCAAGCTGTAAGGCTTCTCCGCGCACCTTTTC 1712  
 Db 552 GAAACAGTGCACCACTGTTAAGGCCCAAGCTGTAAGGCTTCTCCGCGCACCTTTTC 493  
 QY 1713 CGCCCCGGGTGATGCTGCGCAGAGTGGCCCTGTGAGCAATCGGGGAGAGAGAGTGT 1772  
 Db 492 CGCCCCGGGTGATGCTGCGCAGAGTGGCCCTGTGAGCAATCGGGGAGAGAGAGTGT 433  
 QY 1773 GCCTCATGATCGTGTGTCCCGGTGATGATTTCAAGAGTGGCCGTGTGAGACAGAG 1832

Query Match	85.1%;	Score 2033.4;	DB 13;	Length 2177;
Best Local Similarity	96.4%;	Pred. No. 0;		
Matches 2101;	Conservative 0;	Mismatches 76;	Indels 2;	Gaps 2;
QY 11	CGAGCTTGAGGAA	CCAGAGGAAAGGAGAGAAAA	TGGCGTCCAGGATTACAGTACTAT	70
DB 1	GGAGCTTGAGGAA	CCAGAGGAAAGGAGAGAAAA	TGGCATTCAGTAACTTAC	60
QY 71	AGCCAGGTGAGGGG	AGGAGGCTTACAGTGTCTTACA	CCGGCCAGGCCACTCAAGATAT	130
DB 61	AGCCAGGTGAGGGG	AGGAGGCTTACAGTGTCTTACA	CCAGGCCACTCAAGATAT	120
QY 131	GCACAGACCA	CCAGGAGATATGGGCAACAAAGCTATGGA	ACTTATGAGACGCCACCTGAT	190
DB 121	GCACAGACCA	CCAGGAGATATGGGCAACAAAGCTATGGA	ACTTATGAGACGCCACTTGTAT	180
QY 191	GTCAGCTATACCA	GCGCTCAGACCACTGCAACTATGGGAGAGACCGCTTATGCAATTTCT	250	
DB 181	GTCAGCTATACCA	GCGCTCAGACCACTGCAACTATGGGAGAGACCGCTTATGCAATTTCT	240	
QY 251	TATGAGACAGCGCT	CCGCTTACTACTCACTGCACTGCCCCCGAGGATACAGGACGCT	310	
DB 241	TATGAGACAGCGCT	CCGCTTACTACTCACTGCACTGCCCCCGAGGATACAGGACGCT	300	
QY 311	GTCCAGGGGATAT	GGCATCTGCTTATGATATCAACCACTGTCACGTACACCAACCAG	370	
DB 301	GTCCAGGGGATAT	GGCATCTGCTTATGATATCAACCACTGTCACGTACACCAACCAG	360	
QY 371	GCGTCTATGAGG	CTCAGTCTGCAATATGGACCTCAGCGCTTATTCAGACCTATGGGAG	430	
DB 361	GCGTCTATGAGG	CTCAGTCTGCAATATGGACCTCAGCGCTTATTCAGACCTATGGGAG	420	
QY 431	CAGGACAGACGCA	CTGCACTTACAAGACCGGAGATGGAACAAAGGCCACTGAGACCTAGT	490	
DB 421	CAGGACAGACGCA	CTGCACTTACAAGACCGGAGATGGAACAAAGGCCACTGAGACCTAGT	480	
QY 491	CAACCTCAATAT	TACAGAGGGGGTTTACAACAGGCCAGCTAGGATATGAGACAGATTAAC	550	
DB 481	CAACCTCAATAT	TACAGAGGGGGTTTACAACAGGCCAGCTAGGATATGAGACAGATTAAC	540	
QY 551	TACAGTTATCCCA	GAGTACCTGGAGAGTACCCCACTGAGCCAGTCACTGCACTCCATCC	610	
DB 541	TGCAAGTTATCCCA	GAGTACCTGGAGAGTACCCCACTGAGCCAGTCACTGCACTCCATCC	600	
QY 611	TACCTCTCTAC	AGCTATTTCTCTTACACAGCCGACTATGATGATCAGAGAGTTACTCT	670	
DB 601	TACCTCTCTAC	AGCTATTTCTCTTACACAGCCGACTATGATGATCAGAGAGTTACTCT	660	
QY 671	CAGGAGAACCTAT	TGGGCAACGAGAGGCTATGGAACAGAGAGTACTATATGGTCAACAA	730	
DB 661	CAGGAGAACCTAT	TGGGCAACGAGAGGCTATGGAACAGAGAGTACTATATGGTCAACAA	720	
QY 731	AGCAGTATGGGAG	CAGAGCTCCCACTAGTTATCCACCCCAACTGATCTTACAGCCAA	790	



Db 721 AGCAGTATGGGAGCTGCTCCCACTAGTACCACCCCAACTTGATCTACAGCCAA 780  
Qy 791 GCTCCAGTCAATATAGCCCAAGAGAGAGCTACGGGAGAGAGTTCATCCAGC 850  
Db 781 GCTCCAGTCAATATAGCCCAAGAGAGAGCTACGGGAGAGAGTTCATCCAGC 840  
Qy 851 GACCAACCCAGTACATGAGTGTGTTATGAGCAGAGTCTGAGAGATTTTCCGACAGCA 910  
Db 841 GACCAACCCAGTACATGAGTGTGTTATGAGCAGAGTCTGAGAGATTTTCCGACAGCA 900  
Qy 911 GAGAACCGGAGCATGATGAGCCCTGATTAACCGGAGGAGAGAGGAGGATTTGATCGT 970  
Db 901 GAGAACCGGAGCATGATGAGCCCTGATTAACCGGAGGAGAGAGGAGGATTTGATCGT 960  
Qy 971 GAGAGCATGAGCAGAGTGGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1030  
Db 961 GAGAGCATGAGCAGAGTGGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1020  
Qy 1031 CGAGGTGGCTTCAATAGCCCTGGTGAACCCATGATGAAGACAGATCTTGATCTAGGC 1090  
Db 1021 CAAGTTGGCTTCAATAGCCCTGGTGAACCCATGATGAAGACAGATCTTGATCTAGGC 1080  
Qy 1091 CCACTGTAGATCCAGATGAGAGCTGACAAAGTGCATTTATGTACAGAGATTAAT 1150  
Db 1081 CCACTGTAGATCCAGATGAGAGCTGACAAAGTGCATTTATGTACAGAGATTAAT 1140  
Qy 1151 GACAGTGTGATCTAGATGATCTGAGCAGACTCTTTAAGCAGTGTGGGTTGTTAAGATG 1210  
Db 1141 GACAGTGTGATCTAGATGATCTGAGTGAATCTGTTAAGCAGTGTGGGTTGTTAAGATG 1200  
Qy 1211 AACAGAGACTGGGAGCAACCATGATCCATGATCCATCTACCTGAGCAAGAGAAAGAGCC 1270  
Db 1201 AACAGAGACTGGGAGCAACCATGATCCATGATCCATCTACCTGAGCAAGAGAAAGAGCC 1260  
Qy 1271 AAAGCGCATGCCAGATGTCCTATGAAAGCCCACTGCAAGCTGCGTGAATG 1330  
Db 1261 AAAGGTATGTCACAGTGTCTGTGAGAGACTCACTCTGCAAAAGCTGCGTGAATG 1320  
Qy 1331 TTTGATGAGAAAGATTTTCAAGGAGAGCACTTAATCTCTCTGCTCGGAGAGAGCT 1390  
Db 1321 TTTGATGAGAAAGATTTTCAAGGAGAGCACTTAATCTCTCTGCTCGGAGAGAGCT 1380  
Qy 1391 CCAATGAACAGTATGCGGAGGCTGCTGCAACCCGCTGAGGAGAGAGGATGCCACCA 1450  
Db 1381 CCAATGAACAGTATGCGGAGGCTGCTGCAACCCGCTGAGGAGAGAGGATGCCACCA 1440  
Qy 1451 CTCCGTGAGAGTCCAGAGAGCCAGAGAGTCTTGGGAGAGCCATGGTGCATGGAGGC 1510  
Db 1441 CTCTGCGAGAGTCCAGAGAGCCAGAGAGTCTTGGGAGAGCCATGGTGCATGGAGGC 1500  
Qy 1511 CGTGGAGAGATGAGAGAGCTTCTCTCAAGAGAGAGCCCGGGGTTCGAGAGAGAGCC 1570  
Db 1501 CGTGGAGAGATGAGAGAGCTTCTCTCAAGAGAGAGCCCGGGGTTCGAGAGAGAGCC 1560  
Qy 1571 TCTGGAGAGAGAAAGTCCAGAGAGCTGAGAGAGTGGAGAGTCCCAATCCGGGTGT 1630  
Db 1561 TCTGGAGAGAGAAAGTCCAGAGAGCTGAGAGAGTGGAGAGTCCCAATCCGGGTGT 1620  
Qy 1631 GGAAGACAGAACTTCCGCTGAGAGAGAGTGCACCAAGTGTAAAGCCCAAGAGCTGAA 1690  
Db 1621 GGAAGACAGAACTTCCGCTGAGAGAGAGTGCACCAAGTGTAAAGCCCAAGAGCTGAA 1680  
Qy 1691 GGGTTCCTCCCGCAGCCCTTCCGCGCCCGGGGTGTGATGCTGCAAGAGTGCCTGTGT 1750  
Db 1681 GGGTTCCTCCCGCAGCCCTTCCGCGCGGGGTGTGATGCTGCAAGAGTGCCTGTGT 1740  
Qy 1751 GGCATGCGGAGAGAGAGAGAGTGCCTCAATGATGATGATGATGATGATGATGATGAT 1810  
Db 1741 GGCATGCGGAGAGAGAGAGAGTGCCTCAATGATGATGATGATGATGATGATGATGAT 1800  
Qy 1811 GGTGCGCGTGTGAGAGAGAGAGTGTGCTTCCGTGTGTGCGCGGAGCATGAGACGAGTGGC 1870  
Db 1801 GGTGCGTGTGTGAGAGAGAGAGTGTGCTTCCGTGTGTGCGCGGAGCATGAGACGAGTGGC 1859

Qy 1871 TTTGTTGAGAGAGAGAGAGTGGCCCTGGGGGGCCCTGACCTTTGATGAGACAGATG 1930  
Db 1860 TTTGTTGAGAGAGAGAGAGTGGCCCTGGGGGGCCCTGACCTTTGATGAGACAGATG 1918  
Qy 1931 GAG 1990  
Db 1919 GAG 1978  
Qy 1991 GAGCGAGAGATCGGCTTCAATGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2050  
Db 1979 GAGCGAGAGATCGGCTTCAATGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2038  
Qy 2051 TTTATTTTAAACGAGAAATGTTTAAATTTATTTATTTATTTATTTATTTATTTAT 2110  
Db 2039 TTTATTTTAAACGAGAAATGTTTAAATTTATTTATTTATTTATTTATTTATTTAT 2098  
Qy 2111 CAACATTATGATTAATCTGCTGCTGATCTTATGATTTTCAACATTGTTGAAGAAAT 2170  
Db 2099 CAACATTATGATTAATCTGCTGCTGATCTTATGATTTTCAACATTGTTGAAGAAAT 2158  
Qy 2171 TAAACAGATTAATGTTA 2189  
Db 2159 TAAACAGATTAATGTTA 2177

RESULT 12  
AB199383  
ID AB199383 standard; cDNA; 2188 BP.  
XX  
XX AB199383;  
XX  
DT 07-MAR-2002 (first entry)  
DE Mouse ischaemic condition related cDNA sequence SEQ ID NO:288.  
XX  
XX Mouse; Ischaemia; compressive ischaemia; occlusive ischaemia;  
XX  
XX vasospastic ischaemia; ischaemic condition; ischaemic disease; ss.  
OS Mus musculus.  
XX  
XX WO200188188-A2.  
XX  
XX 22-NOV-2001.  
XX  
XX 18-MAY-2001; 2001MO-JP004192.  
XX  
XX 18-MAY-2000; 2000JP-00145977.  
XX  
XX (UNYI-) UNIV NIHON SCHOOL JURIDICAL PERSON.  
XX  
XX Ishikawa K, Arai S, Takahashi Y, Negata T, Ishii Y;  
XX  
XX WPI; 2002-034733/04.  
XX  
XX P-FSDB; ABB57126.  
XX  
XX Examining the ischaemic condition (e.g. occlusive ischaemia) by measuring  
XX  
XX expression levels of particular genes defined in the specification or by  
XX  
XX determining the expression profile of a gene group comprising these  
XX  
XX genes.  
XX  
XX Claim 2; Page 794-799; 2690pp; English.  
XX  
XX The present invention describes a method for examining ischaemic  
XX  
XX conditions, comprising measuring the expression levels of particular  
XX  
XX genes (I) in a test sample or determining the expression profile of a  
XX  
XX gene group in the sample comprising genes selected from (I). The method  
XX  
XX is useful for examining the ischaemic condition (e.g. compressive  
XX  
XX ischaemia, occlusive ischaemia or vasospastic ischaemia) by measuring  
XX  
XX expression levels of particular genes (AB199202 to AB199912, encoding the  
XX  
XX protein sequences in ABB57020 to ABB57374) or by determining the  
XX  
XX expression profile of a gene group comprising these genes. The expression  
XX  
XX levels or expression profiles produced by these genes are used as an





QY 2041 GACTACCAAGTTATTTTAAACGAAATGTTTAAATTATTAATTCATATTATA 2100  
DB 2044 GAGTACCAAGTTATTTTAAACGAAATGTTTAAATTATTAATTCATATTATA 2103  
QY 2101 ATGTGGCCACCAATTATGATTTCTGTCTGACTTATGATTTTTCACCATTTGT 2160  
DB 2104 ATGTGGCCACCAATTATGATTTCTGTCTGACTTATGATTTTTCACCATTTGT 2163  
QY 2161 GAAAGAACATTAAACAGTTAAAT 2185  
DB 2164 GGAGAAACATTAAACAGTTAAAT 2188

RESULT 13  
ADA53506  
ID ADA53506 standard; cDNA; 1988 BP.  
XX  
AC ADA53506;  
XX  
DT 20-NOV-2003 (first entry)  
XX  
DE Human coding sequence, SEQ ID 1074.  
XX  
XX Cytostatic; Anti-inflammatory; Osteopathic; Neuroprotective; Nootropic;  
KM Gene Therapy; human; secretory protein; membrane proteins; cancer;  
KM Inflammatory disease; osteoporosis; neurological disease; gene; ss.  
XX  
OS Homo sapiens.  
XX  
PN BP1293569-A2.  
XX  
PD 19-MAR-2003.  
XX  
PF 21-MAR-2002; 2002EP-00006586.  
XX  
PR 14-SEP-2001; 2001JP-00328381.  
PR 24-JAN-2002; 2002US-0350435P.  
XX  
PA (HELI-) HELIX RES INST.  
PA (REAS-) RES ASSOC BIOTECHNOLOGY.  
XX  
PI Isegai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S;  
PI Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I;  
PI Seki N, Yoshikawa T, Otsuka M, Negahara K, Masuno Y;  
DR WPI; 2003-395539/38.  
DR P-PSDB; ADA55145.  
XX  
XX New polynucleotides encoding full-length polypeptides, e.g. secretory  
PT and/or membrane proteins, useful for developing medicines for diseases in  
PT which the gene is involved, or as target molecules for gene therapy.  
XX  
PS Claim 1; SEQ ID NO 1074; 205pp; English.  
XX  
XX The present invention relates to novel human secretory or membrane  
CC proteins (ADA54072-ADA55710) and their coding sequences (ADA52433-  
CC ADA54071). The coding sequences are useful in the gene therapy of  
CC diseases caused by abnormalities of the proteins, e.g. cancer,  
CC inflammatory diseases, osteoporosis or neurological disease.  
XX  
SQ Sequence 1988 BP; 519 A; 504 C; 584 G; 381 T; 0 U; 0 other;

Query Match 75.7%; Score 1808.4; DB 10; Length 1988;  
Best Local Similarity 92.2%; Pred. No. 0;  
Matches 1987; Conservative 0; Mismatches 1; Indels 168; Gaps 1;

QY 33 AAGAGAGAAATGGCGTCCACGATTAAGTACCTATAGCAAGCTGACGCGAGAG 92  
DB 1 AAGAGAGAAATGGCGTCCACGATTAAGTACCTATAGCAAGCTGACGCGAGAG 60  
QY 93 GCTACAGTGTCTTACCGCCGACCCCACTCAAGGATATGACAGACCCAGCATATG 152

DB 61 GCTACAGTGTCTTACCGCCGACCCCACTCAAGGATATGACAGACCCAGCATATG 120  
QY 153 GCGACAAAGCTATGGAACCTATATGACAGCCCACTGATGCTATACCGAGCTCAGA 212  
DB 121 GCGACAAAGCTATGGAACCTATATGACAGCCCACTGATGCTATACCGAGCTCAGA 180  
QY 213 CCAGTGCACCTATGAGGAGAGCCGCTATGCACTCTTATGAGACAGCCTCCACTGTT 272  
DB 181 CCACTGCACCTATGAGGAGAGCCGCTATGCACTCTTATGAGACAGCCTCCACTGTT 240  
QY 273 ATACTCTCCAACTGCCCCCAGGACCTATGACAGCCAGCCTCTTATGAGCTCAGTGT 332  
DB 241 ATACTCTCCAACTGCCCCCAGGACCTATGACAGCCAGCCTCTTATGAGCTCAGTGT 300  
QY 333 CTATGATACCACTGCTTACAGTCAACCAACCCAGGCTCTTATGAGCTCAGTGT 392  
DB 301 CTATGATACCACTGCTTACAGTCAACCAACCCAGGCTCTTATGAGCTCAGTGT 360  
QY 393 CATATGCACTGAGCTGCTTATCCAGCTTATGAGGAGAGAGAGAGAGAGAGAGAGAG 452  
DB 361 CATATGCACTGAGCTGCTTATCCAGCTTATGAGGAGAGAGAGAGAGAGAGAGAGAG 420  
QY 453 CAAAGCCGAGATGGAACAAAGCCCACTGAGACTAGTCACTCAATCTTACAGAGGG 512  
DB 421 CA-----  
QY 513 GTTACAAACAGCCAGCCTAGATATGACAGAGTAACTTATCCAGAGTACTG 572  
DB 423 -----  
QY 573 GGAGTACCCCATGACAGCAGTCACTGACCTCCATCCTTACAGAGTATTCCT 632  
DB 423 -----AGTATTCCT 432  
QY 633 CTACACAGCCGACTAGTATGATCAAGAGAGTTACTCTACAGAGAAACCTATGGGCAAC 652  
DB 433 CTACACAGCCGACTAGTATGATCAAGAGAGTTACTCTACAGAGAAACCTATGGGCAAC 492  
QY 693 CGAGAGCTATGAG 752  
DB 493 CGAGAGCTATGAG 552  
QY 753 CCACTAGTATACCAACCCCAACTGATCTTACAGAGCCCACTCCAGTCAATATAGCCAC 812  
DB 553 CCACTAGTATACCAACCCCAACTGATCTTACAGAGCCCACTCCAGTCAATATAGCCAC 612  
QY 813 AGAGCAGAGCTACGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 872  
DB 613 AGAGCAGAGCTACGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 672  
QY 873 TTTATGGGAG 932  
DB 673 TTTATGGGAG 732  
QY 933 CTGATTAACCGGAG 992  
DB 733 CTGATTAACCGGAG 792  
QY 993 GGGAG 1052  
DB 793 GGGAG 852  
QY 1053 GTGGAACCATGATGGAAG 1112  
DB 853 GTGGAACCATGATGGAAG 912  
QY 1113 ACTCTGACAAAGTGAATTTATGATCAAGAGATTAATATGACAGTGTGACTTATGATC 1172  
DB 913 ACTCTGACAAAGTGAATTTATGATCAAGAGATTAATATGACAGTGTGACTTATGATC 972  
QY 1173 TGGCAGACTTTTAAAG 1232  
DB 973 TGGCAGACTTTTAAAG 1032

OY	1233	TGATCACATCTACCTGGACAAAGGAAACAGGAAACCCAAAGGCGATGCGCAGTGTCT	1232
Db	1033	TGATCACATCTACCTGGACAAAGGAAACAGGAAACCCAAAGGCGATGCGCAGTGTCT	1092
OY	1233	ATGAAGACCCCACTGCGCAAGCGTCGCTGGAAATGGTTGATGGGAAAGATTTTCAG	1352
Db	1093	ATGAAGACCCCACTGCGCAAGCGTCGCTGGAAATGGTTGATGGGAAAGATTTTCAG	1152
OY	1353	GGAAGCAACTTAAAGTCTCTCTGCTCGGAGAAACCTTCCAATGAACGTATGCGGGGTG	1412
Db	1153	GGAAGCAACTTAAAGTCTCTCTGCTCGGAGAAACCTTCCAATGAACGTATGCGGGGTG	1212
OY	1413	GTCTGCAACCCCGTGAAGGCGAGAGCATGCGACCACTCCTCGTGAAGGTCCAGAGGCG	1472
Db	1213	GTCTGCAACCCCGTGAAGGCGAGAGCATGCGACCACTCCTCGTGAAGGTCCAGAGGCG	1272
OY	1473	CAGAGAGTCTCGGGGGAACCCATGAGTGCATGGAAGGCGCTGGAAGGATAGAGAGGCT	1532
Db	1273	CAGAGAGTCTCGGGGGAACCCATGAGTGCATGGAAGGCGCTGGAAGGATAGAGAGGCT	1332
OY	1533	TCCCTCCAAAGAGAACCCCGGGGTTCCCGAGGGAACCCCTCTGGAAGAGAAACGTCCAGC	1592
Db	1333	TCCCTCCAAAGAGAACCCCGGGGTTCCCGAGGGAACCCCTCTGGAAGAGAAACGTCCAGC	1392
OY	1593	ACCGAGCTGGAGACTGGCAGTGTCTCCCAATCCGGGATTGTGAAACAGAACTTCGCTGGA	1652
Db	1393	ACCGAGCTGGAGACTGGCAGTGTCTCCCAATCCGGGATTGTGAAACAGAACTTCGCTGGA	1452
OY	1653	GAAACAGTGCACCAAGTGTAAAGGCCCAAGCCTTGAAGGCTTCTCCGCGCACCTCTTC	1712
Db	1453	GAAACAGTGCACCAAGTGTAAAGGCCCAAGCCTTGAAGGCTTCTCCGCGCACCTCTTC	1512
OY	1713	CGCCCCCGGGTGTGTATCGTGGCAAGGTGGCCCTGTGTGCAATGCGGGGAGAGAGAGGTG	1772
Db	1513	CGCCCCCGGGTGTGTATCGTGGCAAGGTGGCCCTGTGTGCAATGCGGGGAGAGAGAGGTG	1572
OY	1773	GCCTCATGGAATGATGTGTGGTCCCGGTGGAATGTTGAGAGGTGGCCGTGTGGAAGACAGAG	1832
Db	1573	GCCTCATGGAATGATGTGTGGTCCCGGTGGAATGTTGAGAGGTGGCCGTGTGGAAGACAGAG	1632
OY	1833	GTGGCTTCGTGTGTGGCCCGGGCATGACCGAAGTGGCTTTGTGTGAGAAACGAGGTG	1892
Db	1633	GTGGCTTCGTGTGTGGCCCGGGCATGACCGAAGTGGCTTTGTGTGAGAAACGAGGTG	1692
OY	1893	GCCCTGGGGGGGCCCTCTGCACTTTGATGGAACAGATGGAGAGAAAGAGAGAGAGCTG	1952
Db	1693	GCCCTGGGGGGGCCCTCTGCACTTTGATGGAACAGATGGAGAGAAAGAGAGAGAGCTG	1752
OY	1953	GAGAGCTGGAAAAATGGAATAAGCGCAGACCGTCAGAGGCGCAGAGATCGGCCCTACT	2012
Db	1753	GAGAGCTGGAAAAATGGAATAAGCGCAGACCGTCAGAGGCGCAGAGATCGGCCCTACT	1812
OY	2013	AGATGCAAGAACCCCGCAGAGCTGCATTGACTACACAGATTATTTTAAACAGAAAAAT	2072
Db	1813	AGATGCAAGAACCCCGCAGAGCTGCATTGACTACACAGATTATTTTAAACAGAAAAAT	1872
OY	2073	GTTTTAAATTTATTAATTCATATTTTAAATGTTGGCCCAACATTAATGATTAATTCCTGT	2132
Db	1873	GTTTTAAATTTATTAATTCATATTTTAAATGTTGGCCCAACATTAATGATTAATTCCTGT	1932
OY	2133	CTGTACTTAAATTTTCAACATTTGAGAGAAACATTAATAACAAAGTTAATGGT 2188	
Db	1933	CTGTACTTAAATTTTCAACATTTGAGAGAAACATTAATAACAAAGTTAATGGT 1988	
RESULT 14			
ADRO7446			
ID	ADRO7446 standard; cDNA; 2026 BP.		
XX	ADRO7446;		
DT	04-NOV-2004 (first entry)		

XX	Full length human cDNA useful for treating neurological disease Seq 952.
XX	gene; ss; human; oligo-capping method; diagnostic marker; gene therapy;
KM	osteoporosis; neurological disease; Alzheimer's disease;
KM	Parkinson's disease; dementia; short memory; cancer;
KM	sense or motor function; emotional reaction; fear response; panic;
KW	osteopathic; neuroprotective; nootropic; antiparkinsonian; cytostatic;
KX	tranquilliser.
OS	Homo sapiens.
PN	EP1447413-A2.
PD	18-AUG-2004.
PP	12-FEB-2004; 2004EP-00003145.
PR	14-FEB-2003; 2003JP-00102207.
PR	09-MAY-2003; 2003JP-00131452.
PA	(REAS-) RES ASSOC BIOTECHNOLOGY.
PI	Isoqat T, Yamamoto J, Nishikawa T, Isono Y, Sugiyama T, Otsuki T;
PI	Makamatsu A, Ishii S, Nagai K, Irie R;
DR	WPI; 2004-583265/57.
XX	P-PDSB; ADRO9402.
PT	New 1995 cDNA, useful for treating osteoporosis, neurological diseases,
XX	Alzheimer's diseases, Parkinson's diseases, dementia and various cancers.
PS	Claim 1; SEQ ID NO 952; 2686bp; English.
XX	This invention relates to novel, isolated full length human cDNA
CC	molecules and the encoded proteins thereof. Specifically, it refers to
CC	cDNA clones obtained by an oligo-capping method, where none of these
CC	clones are identical to any known human mRNAs. The present invention
CC	describes an immunoassay to identify agonists and antagonists, as well as
CC	antibodies, antisense molecules and siRNAs that can all be used to bind
CC	to and modulate expression of the cDNA molecules. As such, these
CC	molecules are useful for diagnostic markers or therapeutic targets for
CC	the various diseases or morbid states. In particular, they are useful in
CC	gene therapy for treating osteoporosis, neurological disease, Alzheimer's
CC	disease, Parkinson's disease, dementia, short memory and various cancers,
CC	as well as for maintaining equilibrium of sense or motor function, and
CC	for treating emotional reaction, fear response and panic. Accordingly,
CC	they exhibit osteopathic, neuroprotective, nootropic, antiparkinsonian,
CC	cytostatic and tranquilliser activities. This polynucleotide is a full
CC	length human cDNA sequence of the invention. NOTE: This sequence is not
CC	given in the sequence listing of the specification but can be obtained on
CC	CD-ROM from the European Patent Office, Vienna Sub-office.
XX	
SEQ	Sequence 2026 BP; 540 A; 513 C; 590 G; 383 T; 0 U; 0 Other;
Query Match	74.6%; Score 1782.8; DB 13; Length 2026;
Best Local Similarity	90.9%; Pred. No. 0;
Matches 1984; Conservative	0; Mismatches 42; Indels 156; Gaps 2;
DQ	7 AGACGGAAGCTTGAAGAAGCAAGAGAGAAATGGCGTCCTCAGATTACAGTAC 66
DB	1 AGACGGAAGCTTGAAGAAGCAAGAGAGAAATGGCGTCCTCAGATTACAGTAC 60
DQ	67 CTATAGCCCAAGCTGCAGGCGACGAGGGGTATAGTGCTTACACCGGCCAGCCACTCAAG 126
DB	61 CTATAGCCCAAGCTGCAGGCGACGAGGGGTATAGTGCTTACACCGGCCAGCCACTCAAG 120
DQ	127 ATATGCAACAAGCAACCAGGCAATTTGGGCAACAANGCTATGGAACCTTAGCAAGCCAC 186
DB	121 ATATGCAACAAGCAACCAGGCAATTTGGGCAACAANGCTATGGAACCTTAGCAAGCCAC 180
DQ	187 TGATGTCAAGTATTAACCAAGGCTCAAGACCATGCAACCTTAGGGGCAAGCGGCTTAGAAC 246

Db 181 TGATGAGCTATATCCAGGCTCAGACCACTGCAACCTATGAGGAGACGGCTATATGAAAC 240  
Qy 247 TTTCTTATGACAGCTCTCCACTGTATTACTTCCAACTGCCCCCAGGCAATACAGCA 306  
Db 241 TTTCTTATGACAGCTCTCCACTGTATTACTTCCAACTGCCCCCAGGCAATACAGCA 295  
Qy 307 GCGTATCAGAGGATATGAGCTGTGTATGATACCACTGTACAGTACCAACAC 366  
Db 296 ----- 295  
Qy 367 CAGAGCTCTATGACAGCTGTGATATGAGCACTCAGCTGTATATCAGCTATAGG 426  
Db 296 -----TATTCCTTACACAGCCGACTAGTTATGATACAGAGTTATCTC 339  
Qy 427 GCGACGCGACAGCACTGCACTTACAAACCGGAGATGAAACAAAGCCCACTGAGAC 486  
Db 340 TCGACAGCCAGAGCACTGCACTTACAAACCGGAGATGAAACAAAGCCCACTGAGAC 399  
Qy 487 TAGTCAACCTCAATCTAGCAAGGAGGTTACAACAGCCGAGCTTAGATATGAGACAGAG 546  
Db 400 TAGTCAACCTCAATCTAGCAAGGAGGTTACAACAGCCGAGCTTAGATATGAGACAGAG 459  
Qy 547 TAACTACAGTTATCCCGAGGTAAGTGGAGCTAACCCAGTCAAGCACTGCACTCC 606  
Db 460 TAACTACAGTTATCCCGAGGTAAGTGGAGCTAACCCAGTCAAGCACTGCACTCC 519  
Qy 607 ATCTTACCTCTTACCAAGCTATTCCTTACACAGCCGACTAGTTATGATACAGACAGTTA 666  
Db 520 ATCTTACCTCTTACCAAGCTATTCCTTACACAGCCGACTAGTTATGATACAGACAGTTA 534  
Qy 667 CTCTCAGACAGAACCTATGAGGCAACCGAGCAGCTATGAGCAGAGAGTATGAGTCA 726  
Db 535 -----CAGCTATGAGCAGCAGAGTATGAGTCA 564  
Qy 727 ACAAGAGCTATGAGGAGAGAGCTCTCCACTAGTTACCAACCCCAACCTGATCTTACAG 786  
Db 565 ACAAGAGCTATGAGGAGAGAGCTCTCCACTAGTTACCAACCCCAACCTGATCTTACAG 624  
Qy 787 CCAAGCTCCAAATCAATATAGCCCAACAGACAGAGCTAAGGAGAGAGAGTCAATCCG 846  
Db 625 CCAAGCTCCAAATCAATATAGCCCAACAGACAGAGCTAAGGAGAGAGAGTCAATCCG 684  
Qy 847 ACAAGACCAACCCAGTATGAGGAGAGCTCTCCACTAGTTACCAACCCCAACCTGATCTTACAG 906  
Db 685 ACAAGACCAACCCAGTATGAGGAGAGCTCTCCACTAGTTACCAACCCCAACCTGATCTTACAG 744  
Qy 907 AGAGAGAACCCGAGATGAGTGGCTCTGATTAACCGGAGGAGAGAGAGAGAGAGAGAGAG 966  
Db 745 AGAGAGAACCCGAGATGAGTGGCTCTGATTAACCGGAGGAGAGAGAGAGAGAGAGAGAG 804  
Qy 967 TCGTGAAGGATGAGAGAGAGTGGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1026  
Db 805 TCGTGAAGGATGAGAGAGAGTGGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 864  
Qy 1027 AGAGCGAGTGGCTTCAATTAAGCTGTGAGACCCATGATGAGAGAGAGAGAGAGAGAGAG 1086  
Db 865 AGAGCGAGTGGCTTCAATTAAGCTGTGAGACCCATGATGAGAGAGAGAGAGAGAGAGAGAG 924  
Qy 1087 AGGCGCACCTGTATGATCAAGATGAGAGCTTGAACAAGTGAATTTATATGATCAAGAGAT 1146  
Db 925 AGGCGCACCTGTATGATCAAGATGAGAGCTTGAACAAGTGAATTTATATGATCAAGAGAT 984  
Qy 1147 AATGAGAGTGTATCTAGATGATCTGAGAGAGCTTTTAAAGAGAGTGGGAGTGTAA 1206  
Db 985 AATGAGAGTGTATCTAGATGATCTGAGAGAGCTTTTAAAGAGAGTGGGAGTGTAA 1044  
Qy 1207 GATGAACAAG 1266  
Db 1045 GATGAACAAG 1104  
Qy 1267 GCCCAAG 1326  
Db 1105 GCCCAAG 1164

Qy 1327 ATGCTTGAATGAG 1386  
Db 1165 ATGCTTGAATGAG 1224  
Qy 1387 GCGTCAATGAACAGATATGCGGAG 1446  
Db 1225 GCGTCAATGAACAGATATGCGGAG 1284  
Qy 1447 ACCATCTCGTGAAGAGTCCAG 1506  
Db 1285 ACCATCTCGTGAAGAGTCCAG 1344  
Qy 1507 AGGCGGTGAGAGAGATGAG 1566  
Db 1345 AGGCGGTGAGAGAGATGAG 1404  
Qy 1567 CCCCTCTGAG 1626  
Db 1405 CCCCTCTGAG 1464  
Qy 1627 TTGTGAG 1686  
Db 1465 TTGTGAG 1524  
Qy 1687 TGAAGGCTTCTCCCGGAG 1746  
Db 1525 TGAAGGCTTCTCCCGGAG 1584  
Qy 1747 TGTGTCATATGCGGAG 1806  
Db 1585 TGTGTCATATGCGGAG 1644  
Qy 1807 CAGAGTGGCTGTGTGAG 1866  
Db 1645 CAGAGTGGCTGTGTGAG 1704  
Qy 1867 TGGCTTTGAG 1926  
Db 1705 TGGCTTTGAG 1764  
Qy 1927 GATGAG 1986  
Db 1765 GATGAG 1824  
Qy 1987 TCAAGAGCGCAG 2046  
Db 1825 TCAAGAGCGCAG 1884  
Qy 2047 CAGATTTATTTTAAACAG 2106  
Db 1885 CAGATTTATTTTAAACAG 1944  
Qy 2107 GCCCAACAATATGATATCTCTGTCTGTATGATATTTTCAATTTGTGAAGAA 2166  
Db 1945 GCCCAACAATATGATATCTCTGTCTGTATGATATTTTCAATTTGTGAAGAA 2004  
Qy 2167 ACATTTAAACAGTTTAAATGCT 2188  
Db 2005 ACATTTAAACAGTTTAAATGCT 2026

RESULT 15  
ADP56333  
ID ADP56333 standard; cDNA; 1807 BP.  
XX  
XX ADP56333;  
AC  
XX  
XX 18-NOV-2004 (first entry)  
XX  
DE Human PRO cDNA sequence SEQ ID NO:2309.  
XX  
XX human; PRO; immune related disease; inflammatory immune response;  
KM



QY 1201 TGTAAAGATGACAGAGAACTGGGCAACCATGATCCATCTACTCTGACAGAGAAC 1260  
| | | | |  
Db 994 TGTAAAGATGACAGAGAACTGGGCAACCATGATCCATCTACTCTGACAGAGAAC 1053  
| | | | |  
QY 1261 AGGAAAGCCCAAAAGCGATGCCACATGTCCTATGAAGACCCACCTGCCAAGGCTGC 1320  
| | | | |  
Db 1054 AGGAAAGCCCAAAAGCGATGCCACATGTCCTATGAAGACCCACCTGCCAAGGCTGC 1113  
| | | | |  
QY 1321 CGTGGAAATGTTTGAATGGGAAAGATTTCAAGGGAGCAAACTTAAAGTCTCCCTGTCTCG 1380  
| | | | |  
Db 1114 CGTGGAAATGTTTGAATGGGAAAGATTTCAAGGGAGCAAACTTAAAGTCTCCCTGTCTCG 1173  
| | | | |  
QY 1381 GAAGAAAGCTTCAATGAACAGATGCGGGGTGTCTGCCACCCCGTGAAGGCGACAGGCAAT 1440  
| | | | |  
Db 1174 GAAGAAAGCTTCAATGAACAGATGCGGGGTGTCTGCCACCCCGTGAAGGCGACAGGCAAT 1233  
| | | | |  
QY 1441 GCCACCACTCTCGTGAAGGTCAGAGAGGCCAGAGGTCTTGGGGGACCCATGGGTCTG 1500  
| | | | |  
Db 1234 GCCACCACTCTCGTGAAGGTCAGAGAGGCCAGAGGTCTTGGGGGACCCATGGGTCTG 1293  
| | | | |  
QY 1501 CATGGAGGCGCGTGAAGAGATAGAGAGGCTTCCCTCCAGAGAGACCCCGGGGTCCCG 1560  
| | | | |  
Db 1294 CATGGAGGCGCGTGAAGAGATAGAGAGGCTTCCCTCCAGAGAGACCCCGGGGTCCCG 1353  
| | | | |  
QY 1561 AGGGAACCCCTCTGAGAGAGAAAGTCCAGCAACGAGCTGAGAGCTGGCAGTGTCCCA 1620  
| | | | |  
Db 1354 AGGGAACCCCTCTGAGAGAGAAAGTCCAGCAACGAGCTGAGAGCTGGCAGTGTCCCA 1413  
| | | | |  
QY 1621 TCCGGGTTGTGAAACCAAGAACTTCCCTGAGAAACAGAGTGCAACAGTGTAAAGGCC 1680  
| | | | |  
Db 1414 TCCGGGTTGTGAAACCAAGAACTTCCCTGAGAAACAGAGTGCAACAGTGTAAAGGCC 1473  
| | | | |  
QY 1681 AAAGCCCTGAAGGCTTCTCCCGCCACCCCTTCCGCCCCCGGGTGTGATCGTGGCAGAG 1740  
| | | | |  
Db 1474 AAAGCCCTGAAGGCTTCTCCCGCCACCCCTTCCGCCCCCGGGTGTGATCGTGGCAGAG 1533  
| | | | |  
QY 1741 TGGCCCTGTGTGACATGCGGGAGAGAAAGGTGACCTCATGGAATCGTGTGTCGCCGTGG 1800  
| | | | |  
Db 1534 TGGCCCTGTGTGACATGCGGGAGAGAAAGGTGACCTCATGGAATCGTGTGTCGCCGTGG 1593  
| | | | |  
QY 1801 AATGTTCAAGAGTGGCCGTGTGAGACAGAGGTGGCTTCCGTGTGACCGGGGCATGGA 1860  
| | | | |  
Db 1594 AATGTTCAAGAGTGGCCGTGTGAGACAGAGGTGGCTTCCGTGTGACCGGGGCATGGA 1653  
| | | | |  
QY 1861 CCGAGGTGGCTTTTGTGTGAAGAAACAGAGTGGCCCTGGGGGGCCCCCTTGACCTTTGAT 1920  
| | | | |  
Db 1654 CCGAGGTGGCTTTTGTGTGAAGAAACAGAGTGGCCCTGGGGGGCCCCCTTGACCTTTGAT 1713  
| | | | |  
QY 1921 GGAACAGATGGAGAGAAAGAGAGGACGTGAGAGACTTGAAAAATGATTAAGGCGA 1980  
| | | | |  
Db 1714 GGAACAGATGGAGAGAAAGAGAGGACGTGAGAGACTTGAAAAATGATTAAGGCGA 1773  
| | | | |  
QY 1981 GCACCGTCAAGAGCGCAGAGATCGGCCCTACTAG 2014  
| | | | |  
Db 1774 GCACCGTCAAGAGCGCAGAGATCGGCCCTACTAG 1807  
| | | | |

Search completed: February 20, 2005, 05:30:30  
Job time : 1175.11 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 20, 2005, 04:13:53 / Search time 396.352 Seconds  
(Without alignments)  
9866.749 Million cell updates/sec

Title: US-10-791-017A-1

Perfect score: 2390  
Sequence: 1 agggggagcagcgcgtgag.....taaaaaaaaaaaaaaaaaa 2390

Scoring table: IDENTITY\_NUC  
Gapop 10.0, Gapext 1.0

Searched: 1202784 seqs, 81813359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

#### Database:

Issued Patents NA: \*  
1: /cgn2\_6/ptodata/1/ina/5A COMB.seq: \*  
2: /cgn2\_6/ptodata/1/ina/5B COMB.seq: \*  
3: /cgn2\_6/ptodata/1/ina/6A COMB.seq: \*  
4: /cgn2\_6/ptodata/1/ina/6B COMB.seq: \*  
5: /cgn2\_6/ptodata/1/ina/PTUS COMB.seq: \*  
6: /cgn2\_6/ptodata/1/ina/backfile1.seq: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2369.4	99.1	2371	2 US-08-343-443B-1	Sequence 1, Appl
2	2038.4	85.3	6002	4 US-09-949-016-13696	Sequence 13696, A
3	1556	65.1	1785	4 US-09-949-016-5043	Sequence 5043, Ap
4	1556	65.1	1785	4 US-09-949-016-5044	Sequence 5044, Ap
5	1429.6	59.8	1783	4 US-09-949-016-1954	Sequence 1954, Ap
6	819	34.3	2412	1 US-08-437-027-18	Sequence 18, Appl
7	432.6	18.1	35784	4 US-09-949-016-16785	Sequence 16785, A
8	432.6	18.1	35784	4 US-09-949-016-16786	Sequence 16786, A
9	431	18.0	454	4 US-09-513-899C-1657	Sequence 1657, Ap
10	383.4	16.0	425	2 US-08-343-443B-45	Sequence 45, Appl
11	364.2	15.2	601	4 US-09-949-016-66382	Sequence 66382, A
12	338	14.1	411	4 US-09-621-976-13361	Sequence 13361, A
13	306.2	12.8	333	4 US-09-513-999C-12062	Sequence 12062, A
14	266.4	11.1	954	2 US-08-343-443B-106	Sequence 106, App
15	213	8.9	601	4 US-09-949-016-176641	Sequence 176641, Sequence 176641,
16	213	8.9	601	4 US-09-949-016-17664	Sequence 17664, Sequence 322, App
17	175.2	7.3	1939	4 US-09-919-039-322	Sequence 176637, Sequence 176637,
18	135.8	5.7	601	4 US-09-949-016-176637	Sequence 176637, Sequence 176637,
19	135.8	5.7	601	4 US-09-949-016-176638	Sequence 176638, Sequence 176638,
20	135.8	5.7	601	4 US-09-949-016-176639	Sequence 176639, Sequence 176639,
21	135.8	5.7	601	4 US-09-949-016-176640	Sequence 176640, Sequence 176640,
22	129.4	5.4	601	4 US-09-949-016-176622	Sequence 176622, Sequence 176622,
23	129.4	5.4	601	4 US-09-949-016-176623	Sequence 176623, Sequence 176623,
24	129.4	5.4	601	4 US-09-949-016-176645	Sequence 176645, Sequence 176645,
25	129.4	5.4	601	4 US-09-949-016-176646	Sequence 176646, Sequence 176646,
26	116.4	4.9	450	3 US-09-370-838-145	Sequence 145, App
27	116.4	4.9	450	4 US-09-854-133-145	Sequence 145, App

#### ALIGNMENTS

28	74	3.1	114793	4 US-10-148-806-3	Sequence 3, Appl
29	70.2	2.9	215	4 US-09-513-899C-31510	Sequence 31510, A
30	67	2.8	152132	4 US-09-949-016-13845	Sequence 13845, A
31	67	2.8	152145	4 US-09-949-016-12371	Sequence 12371, A
32	66.4	2.8	311	4 US-09-313-294A-5928	Sequence 5928, A
33	65.6	2.7	7218	1 US-08-232-463-14	Sequence 14, Appl
34	64.6	2.7	601	4 US-09-949-016-176640	Sequence 176640, Sequence 176640,
35	64.6	2.7	601	4 US-09-949-016-176641	Sequence 176641, Sequence 176641,
36	60	2.5	268	1 US-08-437-027-19	Sequence 19, Appl
37	59.8	2.5	1505	1 US-07-915-246-1	Sequence 15, Appl
38	59	2.5	1225	4 US-09-976-594-416	Sequence 416, App
39	58.8	2.5	34230	4 US-09-949-016-12052	Sequence 12052, A
40	58.8	2.5	128470	4 US-09-949-016-13765	Sequence 13765, A
41	58	2.4	3238	2 US-08-343-443B-5	Sequence 5, Appl
42	57.8	2.4	12695	4 US-09-949-016-176775	Sequence 16775, A
43	57.2	2.4	1682	4 US-09-220-132-82	Sequence 82, Appl
44	57.2	2.4	1684	4 US-09-919-039-323	Sequence 323, Appl
45	57	2.4	1926	3 US-09-249-585A-4	Sequence 4, Appl

RESULT 1  
US-08-343-443B-1  
Sequence 1, Application US/08343443B  
Patent No. 5968734  
GENERAL INFORMATION:  
APPLICANT: Aurias, Alain  
APPLICANT: Delattre, Olivier  
APPLICANT: Desmarte, Chantal  
APPLICANT: Melot, Thomas  
APPLICANT: Peter, Martine  
APPLICANT: Ploogastel, Beatrice  
APPLICANT: Thomas, Gilles  
APPLICANT: Zucman, Jessica  
TITLE OF INVENTION: NUCLEIC ACID CORRESPONDING TO A GENE OF  
TITLE OF INVENTION: CHROMOSOME 22 INVOLVED IN RECURRENT CHROMOSOMAL  
TITLE OF INVENTION: TRANSLATIONS ASSOCIATED WITH THE DEVELOPMENT OF CANCEROUS  
TITLE OF INVENTION: TUMORS, AND NUCLEIC ACIDS OF FUSION RESULTING FROM SAID  
TITLE OF INVENTION: TRANSLATIONS  
NUMBER OF SEQUENCES: 129  
CORRESPONDENCE ADDRESS:  
ADDRESSER: Weiser & Associates  
STREET: 230 South Fifteenth Street  
CITY: Philadelphia  
STATE: PA  
COUNTRY: USA  
ZIP: 19102  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: AEDIT 1.0 DOS text editor  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/343,443B  
FILING DATE: 18-NOV-1994  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/FR93/00494  
FILING DATE: 19-MAY-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: FR 92/06123  
FILING DATE: 20-MAY-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Weiser, Gerard J.  
REGISTRATION NUMBER: 19,763  
REFERENCE/DOCKET NUMBER: 989.6121P  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 215-875-8383  
TELEFAX: 215-875-8394  
INFORMATION FOR SEQ. ID NO: 1:  
SEQUENCE CHARACTERISTICS:



Db 1981 GATGGGCTTACTAGATGACAGAGCCCGAGAGCTGATGACTACAGATTATTTT 2040  
Qy 2060 TAAACGAGAAAGTTTAAATTTAATTCATATTATTAATGTTGGCCACACATAT 2119  
Db 2041 TAAACGAGAAAGTTTAAATTTAATTCATATTATTAATGTTGGCCACACATAT 2100  
Qy 2120 GATTATCTTGTCTGTACTTATGATTTTTCACATTTTGTGAAGAAACATTAACAG 2179  
Db 2101 GATTATCTTGTCTGTACTTATGATTTTTCACATTTTGTGAAGAAACATTAACAG 2160  
Qy 2180 TTAATGATGATGCGGAGTTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 2239  
Db 2161 TTAATGATGATGCGGAGTTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 2220  
Qy 2240 TTTAACAATGGAACCCCTGTGAGCATGCTCAGTATCATTTGTGAGAACCAAGAGGCC 2299  
Db 2221 TTTAACAATGGAACCCCTGTGAGCATGCTCAGTATCATTTGTGAGAACCAAGAGGCC 2280  
Qy 2300 TCTTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2359  
Db 2281 TCTTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2340  
Qy 2360 CAAATGTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2390  
Db 2341 CAAATGTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2371

RESULT 2  
US-09-949-016-13696  
Sequence 13696, Application US/09949016  
Percent No. 6812339  
GENERAL INFORMATION:  
APPLICANT: VENTER, J. Craig et al.  
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
FILE REFERENCE: C1001307  
CURRENT APPLICATION NUMBER: US/09/949,016  
PRIOR FILING DATE: 2000-04-14  
PRIOR APPLICATION NUMBER: 60/241,755  
PRIOR FILING DATE: 2000-10-20  
PRIOR APPLICATION NUMBER: 60/237,768  
PRIOR FILING DATE: 2000-10-03  
PRIOR APPLICATION NUMBER: 60/231,498  
PRIOR FILING DATE: 2000-09-08  
NUMBER OF SEQ ID NOS: 207012  
SOFTWARE: FASTSEQ for Windows Version 4.0  
SEQ ID NO 13696  
LENGTH: 6002  
TYPE: DNA  
ORGANISM: Human  
US-09-949-016-13696

Query Match 85.3%; Score 2038.4; DB 4; Length 6002;  
Best Local Similarity 96.4%; Pred. No. 0;  
Matches 2106; Conservative 0; Mismatches 76; Indels 2; Gaps 2;

Qy 6 GAGACGACCTTGAAG 65  
Db 1996 GATGGGACCTTGAAG 2055  
Qy 66 CCTATAGCCAGCTGACGCGACAGAGGCTACAGTGTCTTACACCGCCACGCTCAAG 125  
Db 2056 CCTATAGCCAGCTGACGCGACAGAGGCTACAGTGTCTTACACCGCCACGCTCAAG 2115  
Qy 126 GATATGACAGACACCCAGGACATATGGGCAAAAGCTATGGAACCTATGACAGCCCA 185  
Db 2116 GATATGACAGACACCCAGGACATATGGGCAAAAGCTATGGAACCTATGACAGCCCA 2175  
Qy 186 CTGATGCTAGCTATACCCAGGCTCAGACCACTGCAACCTATGGGACAGACCGCTATGCA 245  
Db 2176 TTGATGCTAGCTATACCCAGGCTCAGACCACTGCAACCTATGGGACAGACCGCTATGCA 2235

Qy 246 CTTCTATGAGACAGCCCTCCACATGTTATTAATCTCCAACTGCCCCCGAGGCAATACGCC 305  
Db 2236 CTTCTATGAGACAGCCCTCCACATGTTATTAATCTCCAACTGCCCCCGAGGCAATACGCC 2295  
Qy 306 AGCCTGTCAGGGGTATGCACTGCTGCTATGATATACCACTGCTACATGCAACCA 365  
Db 2296 AGCCTGTCAGGGGTATGCACTGCTGCTATGATATACCACTGCTACATGCAACCA 2355  
Qy 366 CCGAGGCTCTATGACAGTCACTGATATGAGCACTGACCTGCTTATTCAGGCTATG 425  
Db 2356 CCGAGGCTCTATGACAGTCACTGATATGAGCACTGACCTGCTTATTCAGGCTATG 2415  
Qy 426 GGGAGAGCCAGACGACCTGACCTACCAAGACCGGAGATGGAACCAACCCCACTGAG 485  
Db 2416 GGGAGAGCCAGACGACCTGACCTACCAAGACCGGAGATGGAACCAACCCCACTGAG 2475  
Qy 486 CTAGTCAACTCAATCTAGACAGGGGGTTAACAACGACCGGCTAGATATGAGACAGA 545  
Db 2476 CTAGTCAACTCAATCTAGACAGGGGGTTAACAACGACCGGCTAGATATGAGACAGA 2535  
Qy 546 GTAACCTACAGTATCCCGAGGTAACCTGAGAGCTACCCCATGACGACGACCTC 605  
Db 2536 GTAACCTACAGTATCCCGAGGTAACCTGAGAGCTACCCCATGACGACGACCTC 2595  
Qy 606 CATCTACCTCTTACAGGATATTTCTCTACACAGCCGATAGTATGATCAGAGAGTT 665  
Db 2596 CATCTACCTCTTACAGGATATTTCTCTACACAGCCGATAGTATGATCAGAGAGTT 2655  
Qy 666 ACTCTGACAGACACCTATGAGGCAACCGAGAGCTATGAGACAGAGAGTATGATGCT 725  
Db 2656 ACTCTGACAGACACCTATGAGGCAACCGAGAGCTATGAGACAGAGAGTATGATGCT 2715  
Qy 726 AACAAAGCAGCTATGAGGACAGCCTCCCACTAGTTAACCAACCCAACTGATCTTACA 785  
Db 2716 AACAAAGCAGCTATGAGGACAGCCTCCCACTAGTTAACCAACCCAACTGATCTTACA 2775  
Qy 786 GCCAAGCTCCATGATCAATATAGCCACAGAGAGAGCTTCCGACAGAGATTCATTC 845  
Db 2776 GCCAAGCTCCATGATCAATATAGCCACAGAGAGAGCTTCCGACAGAGATTCATTC 2835  
Qy 846 GACAGGACCAACCCAGTACGATGAGTATGAGGAGAGAGTCTGAGAGATTTCCGAC 905  
Db 2836 GACAGGACCAACCCAGTACGATGAGTATGAGGAGAGAGTCTGAGAGATTTCCGAC 2895  
Qy 906 CAGAGAGAAACCGAGCATGAGTGGCCCTGATTAACCGGGGACAGGGAAGAGGGGATTTG 965  
Db 2896 CAGAGAGAAACCGAGCATGAGTGGCCCTGATTAACCGGGGACAGGGAAGAGGGGATTTG 2955  
Qy 966 ATCTGAGAGCATGAGCAGAGTGGGCGGGGAGAGAGAGCGCGGTGATGAGGCGCTG 1025  
Db 2956 ATCTGAGAGCATGAGCAGAGTGGGCGGGGAGAGAGAGCGCGGTGATGAGGCGCTG 3015  
Qy 1026 GAGAGGAGGTGCTTCAATTAACCTGATGAGACCCATGAGATGAGAGGACGAGCTTGATC 1085  
Db 3016 GAGAGGAGGTGCTTCAATTAACCTGATGAGACCCATGAGATGAGAGGACGAGCTTGATC 3075  
Qy 1086 TAGGCCACCTGATGATGAGATGAGAGAGCTGCAACAGTGAATTTATGACAGAGAT 1145  
Db 3076 TAGGCCACCTGATGATGAGATGAGAGAGCTGCAACAGTGAATTTATGACAGAGAT 3135  
Qy 1146 TAAATGACAGTGTGATCTTATGATGATGAGAGCTTTCTTTAAGCAGTGTGGGTTTGA 1205  
Db 3136 TAAATGACAGTGTGATCTTATGATGATGAGAGCTTTCTTTAAGCAGTGTGGGTTTGA 3195  
Qy 1206 AGATGAACAAGAGAACTGGGCAACCCATGATGACATTAACCTGAGCAAGAGAAACAGAA 1265  
Db 3196 AGATGAACAAGAGAACTGGGCAACCCATGATGACATTAACCTGAGCAAGAGAAACAGAA 3255  
Qy 1266 AGCCCAAAGGCGATGACAGTGTCTTATGAGAGACCAACCTGAGCAAGGCTGCGTGG 1325  
Db 3256 AGCCCAAAGGCGATGACAGTGTCTTATGAGAGCTCACTATGCAAGGCTGCGTGG 3315  
Qy 1326 AATGCTTATGAGGAGAAAGTTTCAAGGAGCAAACTTAAGTCTCTTCTGCGAGAGA 1385

```

Db      3316  AATGTTTATGGAGAAATTTTCAAGAGCAAACTTAATGCTCTGCTCGAGAGA 3375
Qy      1386  AGCCTCCATGAAACATGATCGGGGTGTCTGCCACCCGTGAGGAGAGGATGCCAC 1445
Db      3376  GGCTCCAGTGAACAGTATCAAGGTGTATGCCACCCCGTGAAGGAGAGGATGCCAC 1435
Qy      1446  CACCACTTCCTGAGAGTTCAGAGAGGCCAGAGGTCTTGGGGGACCCATGGGTTCGATG 1505
Db      3436  CACCACTTCCTGAGAGTTCAGAGAGGCCAGAGGTCTTGGGGGACCCATGGGTTCGATG 1495
Qy      1506  GAGGCGGTGAGAGATAGAGAGGCTTCCCTCCAAAGAGACCCCGGGGTTCCCGAGGA 1565
Db      3496  GAGGCGGTGAGAGATAGAGAGGCTTCCCTCCAAAGAGACCCCGGGGTTCCCGAGGA 1555
Qy      1566  ACCCTCTGAGAGAGAGAACTGTCAGACAGGCTGAGAGCTGGCACTGTTCCAAATCCG 1625
Db      3556  ACACCTCTGAGAGAGAGAACTGTCAGACAGGCTGAGAGCAAGTGTAAAGCTTCAAGC 1615
Qy      1626  GTTGTGAAACCAAACTTGGCTGAGAAACAGATGCCAACAGTGTAAAGCTTCAAGC 1685
Db      3616  GTTGTGAAACCAAACTTGGCTGAGAAACAGATGCCAACAGTGTAAAGCTTCAAGC 1675
Qy      1686  CTGAAGGCTTCTCCCGCACCCTTTCGCGCCCGGGTGTGATCTGAGAGAGTGGCC 1745
Db      3676  CTGAAGGCTTCTCCCGCACCCTTTCGCGCCCGGGTGTGATCTGAGAGAGTGGCC 1735
Qy      1746  CTGGTGGCATGCGGGAGAGAAAGAGTGGCTCATGATCTGTGTGTCTCCGTTGAAATGT 1805
Db      3736  CTGGTGGCATGCGGGAGAGAAAGAGTGGCTCATGATCTGTGTGTCTCCGTTGAAATGT 1795
Qy      1806  TCAGAGTGGCCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1865
Db      3796  TCAGAGTGGCCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1854
Qy      1866  GTGCTTTGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1925
Db      3855  GTGCTTTGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1913
Qy      1926  AGATGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1985
Db      3914  CAATGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1973
Qy      1986  GTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2045
Db      3974  GTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2033
Qy      2046  CCAGATTTATTTTAAACAGAAATGTTTAAATTTAAATTTAAATTTAAATTTAAATGT 2105
Db      4034  CCAGATTTATTTTAAACAGAAATGTTTAAATTTAAATTTAAATTTAAATTTAAATGT 2093
Qy      2106  GGCCCAACATTTATTTATTTCTGTCTGTACTTTAGATTTTTCACCATTTGTGAGA 2165
Db      4094  GGCCCAACATTTATTTATTTCTGTCTGTACTTTAGATTTTTCACCATTTGTGAGA 2153
Qy      2166  AACATTTAAACAGTTAAATGTA 2189
Db      4154  AACATTTAAACAGTTAAATGTA 4177

```

```

; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 5043
; LENGTH: 1785
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-5043

Query Match      65.1%; Score 1556; DB 4; Length 1785;
Best Local Similarity 89.1%; Pred. No. 0;
Matches 1785; Conservative 0; Mismatches 0; Indels 219; Gaps 1;

Qy      11  GGAAGTTGAGAGAACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 70
Db      1  GGAAGTTGAGAGAACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 60
Qy      71  AGCCAAAGCTGACAGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 130
Db      61  AGCCAAAGCTGACAGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 120
Qy      131  GCAAGACCAACAGGAGATATGAGGCAACAAAGTATGAACTATGAGAGAGAGAGAGAGAG 190
Db      121  GCAAGACCAACAGGAGATATGAGGCAACAAAGTATGAACTATGAGAGAGAGAGAGAGAG 180
Qy      191  GTGAGTATACCAAGGCTGAGACCACTGCAACTATGAGAGAGAGAGAGAGAGAGAGAGAG 250
Db      181  GTGAGTATACCAAGGCTGAGACCACTGCAACTATGAGAGAGAGAGAGAGAGAGAGAGAG 240
Qy      251  TATGAGCAAGCTCCCACTGTGTATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 310
Db      241  TATGAGCAAGCTCCCACTGTGTATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 300
Qy      311  GTCCAGAGGATATGAGCACTGTGTGTATTAATTAATTAATTAATTAATTAATTAATTAAT 370
Db      301  GTCCAGAGGATATGAGCACTGTGTGTATTAATTAATTAATTAATTAATTAATTAATTAAT 360
Qy      371  GCCTCTTATGACAGCTGATGTGTATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 430
Db      361  GCCTCTTATGACAGCTGATGTGTATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 420
Qy      431  CAGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 490
Db      421  CAGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 480
Qy      491  CAACCTCAATCTAGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 550
Db      481  CAACCTCAATCTAGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 540
Qy      551  TACAGTTATCCAGAGTACCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 610
Db      541  TACAGTTATCCAGAGTACCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 600
Qy      611  TACCTCTTACAGAGTATCTTCTTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 670
Db      601  TACCTCTTACAGAGTATCTTCTTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 660
Qy      671  CAGCAGAACCATATGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 730
Db      661  CAGCAGAACCATATGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 720
Qy      731  AGCAGTATGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 790
Db      721  AGCAGTATGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 780
Qy      791  GCTCCAGTAAATATGAGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 850
Db      781  GCTCCAGTAAATATGAGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 826
Qy      851  GACACCCAGTATGAGTGTGTATGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 910

```

```

RESULT 3
US-09-949-016-5043
; Sequence 5043, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: C0001307
; CURRENT APPLICATION NUMBER: US/09/949, 016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20

```

[illegible]

DB	1702	GAGGAGAAGAGAGAGACGTGGAGACCTGGAAAAATGATATAAAGCGACACCGTCAAG	1761
OY	1991	GAGCGCAGAGATCGGCCCTACTAG	2014
Db	1762	GAGCGCAGAGATCGGCCCTACTAG	1785
 RESULT 4 US-09-949-016-5044 ; Sequence 5044, Application US/09949016 ; Patent No. 6812339 ; GENERAL INFORMATION: ; APPLICANT: VENTER, J. Craig et al. ; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED ; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF ; FILE REFERENCE: CLO01307 ; CURRENT APPLICATION NUMBER: US/09/949,016 ; PRIORITY FILING DATE: 2000-04-14 ; PRIOR APPLICATION NUMBER: 60/241,755 ; PRIOR FILING DATE: 2000-10-20 ; PRIOR APPLICATION NUMBER: 60/237,768 ; PRIOR FILING DATE: 2000-10-03 ; PRIOR APPLICATION NUMBER: 60/231,498 ; PRIOR FILING DATE: 2000-09-08 ; NUMBER OF SEQ ID NOS: 207012 ; SOFTWARE: FastSeq for Windows Version 4.0 ; SEQ ID NO 5044 ; LENGTH: 1785 ; TYPE: DNA ; ORGANISM: Human US-09-949-016-5044			
 Query Match                65.1%; Score 1556; DB 4; Length 1785; Best Local Similarity     89.1%; Pred. No. 0; Matches 1785; Conservative   0; Mismatches   0; Indels   219; Gaps   1;			
OY	11	GCAGCTTAGAGAAAGAGAGAGAGAGAAAATGGCGTCACGAGTTACGATCTAT	70
Db	1	GGAGCTTGAGGAACGAGAGAGAGAGAAAATGGCGTCACGAGTTACGATCTAT	60
OY	71	AGCCAGGTGAGGCGCAGAGGCGCTACAGTGGCTTACCGCCACCCCATAGGATAT	130
Db	61	AGCCAGGTGAGGCGCAGAGGCGCTACAGTGGCTTACCGCCACCCCATAGGATAT	120
OY	131	GCACAGACCACCCAGGSCATATGGGCAACAAGCTATGGAACCTATGACAGCCCAT	190
Db	121	GCACAGACCACCCAGGSCATATGGGCAACAAGCTATGGAACCTATGACAGCCCAT	180
OY	191	GTCAGCTATACCCAGGCTCAAGCACCTGCAACCTATGGGCGACAGCCGCTTAATCT	250
Db	181	GTCAGCTATACCCAGGCTCAAGCACCTGCAACCTATGGGCGACAGCCGCTTAATCT	240
OY	251	TATGACAGCCTCCCATGCTTATATCTCAATGCCCCCAGGACATACAGCCAGCT	310
Db	241	TATGACAGCCTCCCATGCTTATATCTCAATGCCCCCAGGACATACAGCCAGCT	300
OY	311	GTCCAGGGGATTGGCAGTGTGCTTATATACCACTGCTACAGTCAACCAACCCAG	370
Db	301	GTCCAGGGGATTGGCAGTGTGCTTATATACCACTGCTACAGTCAACCAACCCAG	360
OY	371	GCTCTCTATGAGCTCAGTCTGTCATATGGCACTCAGCCTGCTTATCCAGCCTATG	430
Db	361	GCTCTCTATGAGCTCAGTCTGTCATATGGCACTCAGCCTGCTTATCCAGCCTATG	420
OY	431	CAGCGCAGACCTGCACTGCACTTACAAAGACCGCAGATGGAACAAGCCCATGAGCT	490
Db	421	CAGCGCAGACCTGCACTGCACTTACAAAGACCGCAGATGGAACAAGCCCATGAGCT	480
OY	491	CAACCTCAATCTAGACACAGGGGGTTAACAAACAGACCCAGCTAGATATGACAGATAC	550
Db	481	CAACCTCAATCTAGACACAGGGGGTTAACAAACAGACCCAGCTAGATATGACAGATAC	540
OY	551	TACGATTATCCCAGGTTACTGGGAGTACCCCATGACGACGATCACTGCATCTTCACC	610

```

Db      541 TACGATTTATCCCAAGGATCTGGAGGATACCCCAATGACCACTGACCTCCATCC 600
Qy      611 TACCTCTTACGAGCTATCTCTTACACAGCCCACTAGTTATGATCAGACAGTACTCT 670
Db      601 TACCTCTTACGAGCTATCTCTTACACAGCCCACTAGTTATGATCAGACAGTACTCT 660
Qy      671 CAGAGAAACCTATGGGCAACCGAGAGCTATGAGAGAGAGAGTATGATGATGATGAT 730
Db      661 CAGAGAAACCTATGGGCAACCGAGAGCTATGAGAGAGAGAGTATGATGATGATGAT 720
Qy      731 AGCAGTATGGGCAAGCAGCTCCCACTAGTTACCCCACTGATCTTACAGCCAA 790
Db      721 AGCAGTATGGGCAAGCAGCTCCCACTAGTTACCCCACTGATCTTACAGCCAA 780
Qy      791 GCTCCAGTCAATATGAGCAACAGAGCAAGCTACGGGCAAGAGTTCATCCAGAG 850
Db      781 GCTCCAGTCAATATGAGCAACAGAGCAAGCTACGGGCAAGAGTTCATCCAGAG 826
Qy      851 GACCAACCCAGTACATGGGCTTTATGGGCAAGAGTCTGAGAGATTTTCCGACAGAA 910
Db      827 ----- 826
Qy      911 GAGAACCGAGCATGAGTGGCTCTGATTAACCGGGCAAGGAGAGGGGATTTGATCGT 970
Db      827 ----- 826
Qy      971 GAGAGCATGAGCAGAGTGGGCGGGGAGAGAGAGCGCGGTGATGAGGACGCTGAGAG 1030
Db      827 ----- 826
Qy      1031 CGAGTGGCTTCAATTAAGCTGTGGACCCATGAGTGAAGGACCACTTTGATCTTATG 1090
Db      827 ----- 826
Qy      1091 CCACTGTAGTCCAGATGAGATCTCTGACAAAGTGCATTTATGATCAAGATTAAT 1150
Db      862 CCACTGTAGTCCAGATGAGATCTCTGACAAAGTGCATTTATGATCAAGATTAAT 921
Qy      1151 GACAGTGTAGTCTGATGATGATCTGAGAGCTTTTAAAGAGTGTGGGCTTTAAAGAT 1210
Db      922 GACAGTGTAGTCTGATGATGATCTGAGAGCTTTTAAAGAGTGTGGGCTTTAAAGAT 981
Qy      1211 AACAGAGATCTGGGCAACCCATGATCACTCACTGAGCAAGAGAAACGAGAAAGCC 1270
Db      982 AACAGAGATCTGGGCAACCCATGATCACTCACTGAGCAAGAGAAACGAGAAAGCC 1041
Qy      1271 AAAGGCAATGCAAGTGTCTCTATGAAGACCCCACTGCAAGAGTGTGGGAAATG 1330
Db      1042 AAAGGCAATGCAAGTGTCTCTATGAAGACCCCACTGCAAGAGTGTGGGAAATG 1101
Qy      1331 TTTGATGGGAAAGATTTTCAAGGAGCAAACTTAAAGTCTCTGCTCGAAGAAAGCT 1390
Db      1102 TTTGATGGGAAAGATTTTCAAGGAGCAAACTTAAAGTCTCTGCTCGAAGAAAGCT 1161
Qy      1391 CCAATGAACAGTATGCGGGGTGTCTGCAACCCGCTGAGGAGAGGAGATCCACCA 1450
Db      1162 CCAATGAACAGTATGCGGGGTGTCTGCAACCCGCTGAGGAGAGGAGATCCACCA 1221
Qy      1451 CTCCGTGAGGTTCAGAGAGCCCAAGAGTCTTGGGAGACCATGGTGTGATGAGAGC 1510
Db      1222 CTCCGTGAGGTTCAGAGAGCCCAAGAGTCTTGGGAGACCATGGTGTGATGAGAGC 1281
Qy      1511 CTTGAGAGAGATTAAGAGAGCTTCTCTCAAGAGAGCCCGGGGTTCCTCGAGGAAAGCC 1570
Db      1282 CTTGAGAGAGATTAAGAGAGCTTCTCTCAAGAGAGCCCGGGGTTCCTCGAGGAAAGCC 1341
Qy      1571 TCTGAGAGAGAGAGAGTCTCAGACAGAGCTGAGAGTGGGAGTCCCAATCGGGGTGT 1630
Db      1342 TCTGAGAGAGAGAGAGTCTCAGACAGAGCTGAGAGTGGGAGTCCCAATCGGGGTGT 1401
Qy      1631 GGAAGACAGAACTTGTGCTGAGAGAGAGAGTGAACAGAGTGAAGGCCCCAAGCCTGAA 1690

```

```

Db      1402 GGAAGACAGAACTTGTGCTGAGAAAGAGTGAACCAAGTGAAGGCCCCAAGCCTGAA 1461
Qy      1691 GGTCTTCTCCCGCACCTTTTCCGCCCCGGGTGTGATCTGTGAGAGAGTGGCCCTGT 1750
Db      1462 GGTCTTCTCCCGCACCTTTTCCGCCCCGGGTGTGATCTGTGAGAGAGTGGCCCTGT 1521
Qy      1751 GGCATGCGGGGAGAGAGAGTGGCTCATGATCTGTGTGTGTCCCGGTGAATGTTGAGA 1810
Db      1522 GGCATGCGGGGAGAGAGAGTGGCTCATGATCTGTGTGTGTCCCGGTGAATGTTGAGA 1581
Qy      1811 GGTGCGGTGTGTGAGACAGAGTGGCTTCCGTGTGTGTGCGCGGAGCATGACAGAGTGC 1870
Db      1582 GGTGCGGTGTGTGAGACAGAGTGGCTTCCGTGTGTGTGTGCGCGGAGCATGACAGAGTGC 1641
Qy      1871 TTTGTGTGAGAGAGAGAGTGGCCCTGTGGGGGGCCCCCTGGAACCTTGAATGGAAGATG 1930
Db      1642 TTTGTGTGAGAGAGAGAGTGGCCCTGTGGGGGGCCCCCTGGAACCTTGAATGGAAGATG 1701
Qy      1931 GAGGAGAGAGAGAGAGAGTGGAGACCTGGAATAATGATTAAGGCGAGCACCGTCA 1990
Db      1702 GAGGAGAGAGAGAGAGAGTGGAGACCTGGAATAATGATTAAGGCGAGCACCGTCA 1761
Qy      1991 GAGCGAGAGAGTGGCTCTACTG 2014
Db      1762 GAGCGAGAGAGTGGCTCTACTG 1785

RESULT 5
US-09-949-016-1954
; Sequence 1954, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1954
; LENGTH: 1783
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-1954

Query Match      59.8%; Score 1429.6; DB 4; Length 1783;
Best Local Similarity 85.8%; Pred. No. 0;
Matches 1719; Conservative 0; Mismatches 64; Indels 221; Gaps 3;

```



Db 241 TATGAGCAGCCTCCCACTGGTTATACCTCCAACTGCCCCCAAGGATACAGCAGCCT 300  
 Qy 311 GTCCAGGGGATATGGCACTGGTCTTATGATACCACTGCTACAGTCAACCACTGAG 370  
 Db 301 GTCCAGGGGATATGGCACTGGTCTTATGATACCACTGCTACAGTCAACCACTGAG 360  
 Qy 371 GCGCTCTATGAGCTGATGCTGATGATGAGCTGAGCTGCTTATGAGCTTATGAGGAG 430  
 Db 361 GCGCTCTATGAGCTGATGCTGATGATGAGCTGAGCTGCTTATGAGCTTATGAGGAG 420  
 Qy 431 CAGCAGCAGCAGCTGAGCTGATGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAG 490  
 Db 421 CAGCAGCAGCAGCTGAGCTGATGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAG 480  
 Qy 491 CAACTCTATGATGAGCAGAGGGGTTTAAACAGCAGCTGAGCTGAGCTGAGCTGAG 550  
 Db 481 CAACTCTATGATGAGCAGAGGGGTTTAAACAGCAGCTGAGCTGAGCTGAGCTGAG 540  
 Qy 551 TACAGTATCCCAAGTACCTGAGAGCTACCCCACTGAGCAGCTGAGCTGAGCTGAG 610  
 Db 541 TACAGTATCCCAAGTACCTGAGAGCTACCCCACTGAGCAGCTGAGCTGAGCTGAG 600  
 Qy 611 TACCTCTCTACAGCTATCTCTCTACAGCAGCTGAGCTGAGCTGAGCTGAGCTGAG 670  
 Db 601 TACCTCTCTACAGCTATCTCTCTACAGCAGCTGAGCTGAGCTGAGCTGAGCTGAG 660  
 Qy 671 CAGCAGAACTATGAGGAGCAGCAGCAGCTGAGCAGCAGCTGAGCTGAGCTGAG 730  
 Db 661 CAGCAGAACTATGAGGAGCAGCAGCAGCTGAGCAGCAGCTGAGCTGAGCTGAG 720  
 Qy 731 AGCAGTATGAGGAGCAGCAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAG 790  
 Db 721 AGCAGTATGAGGAGCAGCAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAG 780  
 Qy 791 GCTCAGTATGAGTATGAGCAGCAGCAGCTGAGCAGCAGCTGAGCTGAGCTGAG 850  
 Db 781 GCTCAGTATGAGTATGAGCAGCAGCAGCTGAGCAGCAGCTGAGCTGAGCTGAG 826  
 Qy 851 GACCACTCAGTATGAGTATGAGTATGAGCAGCAGCTGAGCAGCTGAGCTGAG 910  
 Db 827 ----- 826  
 Qy 911 GAGAACCGAGCAGTATGAGTATGAGTATGAGCAGCAGCTGAGCTGAGCTGAG 970  
 Db 827 ----- 826  
 Qy 971 GAGAGCAGTATGAGTATGAGTATGAGTATGAGCAGCAGCTGAGCTGAGCTGAG 1030  
 Db 827 ----- 826  
 Qy 1031 CGAGTATGAGTATGAGTATGAGTATGAGCAGCAGCTGAGCTGAGCTGAGCTGAG 1090  
 Db 827 -----GACCATGAGTATGAGTATGAGTATGAGCAGCAGCTGAGCTGAGCTGAG 861  
 Qy 1091 CCACTCTATGAGTATGAGTATGAGTATGAGCAGCAGCTGAGCTGAGCTGAGCTGAG 1150  
 Db 862 CCACTCTATGAGTATGAGTATGAGTATGAGCAGCAGCTGAGCTGAGCTGAGCTGAG 921  
 Qy 1151 GACAGTATGAGTATGAGTATGAGTATGAGCAGCAGCTGAGCTGAGCTGAGCTGAG 1210  
 Db 922 GACAGTATGAGTATGAGTATGAGTATGAGCAGCAGCTGAGCTGAGCTGAGCTGAG 981  
 Qy 1211 AACAGAGAACTGAGCAGCAGCTGAGTATGAGTATGAGCAGCAGCTGAGCTGAG 1270  
 Db 982 AACAGAGAACTGAGCAGCAGCTGAGTATGAGTATGAGCAGCAGCTGAGCTGAG 1041  
 Qy 1271 AAAGGAGTATGAGTATGAGTATGAGTATGAGCAGCAGCTGAGCTGAGCTGAG 1330  
 Db 1042 AAAGGAGTATGAGTATGAGTATGAGTATGAGCAGCAGCTGAGCTGAGCTGAG 1101  
 Qy 1331 TTGATGAGGAAAGTTTAAAGGAGAACTTAAAGTATGAGTATGAGCAGCAGCT 1390

Db 1102 TTGATGAGGAAAGTTTAAAGGAGAACTTAAAGTATGAGTATGAGCAGCAGCT 1161  
 Qy 1391 CCAATGAACTATGAGGAGTATGAGTATGAGCAGCAGCTGAGTATGAGCAGCAGCT 1450  
 Db 1162 CCAATGAACTATGAGGAGTATGAGTATGAGCAGCAGCTGAGTATGAGCAGCAGCT 1221  
 Qy 1451 CCGAGTATGAGTATGAGGAGTATGAGTATGAGCAGCAGCTGAGTATGAGCAGCAGCT 1510  
 Db 1222 CCGAGTATGAGTATGAGGAGTATGAGTATGAGCAGCAGCTGAGTATGAGCAGCAGCT 1281  
 Qy 1511 CCGAGTATGAGTATGAGGAGTATGAGTATGAGCAGCAGCTGAGTATGAGCAGCAGCT 1570  
 Db 1282 CCGAGTATGAGTATGAGGAGTATGAGTATGAGCAGCAGCTGAGTATGAGCAGCAGCT 1341  
 Qy 1571 TCGAGAGAGAGAACTGAGCAGCAGCTGAGTATGAGTATGAGCAGCAGCTGAGTATGAG 1630  
 Db 1342 TCGAGAGAGAGAACTGAGCAGCAGCTGAGTATGAGTATGAGCAGCAGCTGAGTATGAG 1401  
 Qy 1631 GGAAGCAGAACTGAGCAGCAGCTGAGTATGAGTATGAGCAGCAGCTGAGTATGAG 1690  
 Db 1402 GGAAGCAGAACTGAGCAGCAGCTGAGTATGAGTATGAGCAGCAGCTGAGTATGAG 1461  
 Qy 1691 GCGTCTCTCCGAGCAGCAGCTGAGTATGAGTATGAGCAGCAGCTGAGTATGAG 1750  
 Db 1462 GCGTCTCTCCGAGCAGCAGCTGAGTATGAGTATGAGCAGCAGCTGAGTATGAG 1521  
 Qy 1751 GCGATGAGGAGAGAGAGTATGAGTATGAGCAGCAGCTGAGTATGAGCAGCAGCTGAG 1810  
 Db 1522 GCGATGAGGAGAGAGAGTATGAGTATGAGCAGCAGCTGAGTATGAGCAGCAGCTGAG 1581  
 Qy 1811 GCGATGAGGAGAGAGAGTATGAGTATGAGCAGCAGCTGAGTATGAGCAGCAGCTGAG 1870  
 Db 1582 GCGATGAGGAGAGAGAGTATGAGTATGAGCAGCAGCTGAGTATGAGCAGCAGCTGAG 1640  
 Qy 1871 TTTGATGAGGAGAGAGAGTATGAGTATGAGCAGCAGCTGAGTATGAGCAGCAGCTGAG 1930  
 Db 1641 TTTGATGAGGAGAGAGAGTATGAGTATGAGCAGCAGCTGAGTATGAGCAGCAGCTGAG 1699  
 Qy 1931 GAGAGAGAGAGAGAGAGTATGAGTATGAGCAGCAGCTGAGTATGAGCAGCAGCTGAG 1990  
 Db 1700 GAGAGAGAGAGAGAGAGTATGAGTATGAGCAGCAGCTGAGTATGAGCAGCAGCTGAG 1759  
 Qy 1991 GAGAGAGAGAGAGAGTATGAGTATGAGCAGCAGCTGAGTATGAGCAGCAGCTGAG 2014  
 Db 1760 GAGAGAGAGAGAGAGTATGAGTATGAGCAGCAGCTGAGTATGAGCAGCAGCTGAG 1783

RESULT 6  
 US-08-437-027-18  
 ? Sequence 18, Application US/08437027  
 ? Patent No. 5670317  
 ? GENERAL INFORMATION:  
 ? APPLICANT: Landany1, Marc  
 ? APPLICANT: Gerald, William  
 ? TITLE OF INVENTION: A DIAGNOSTIC TEST FOR TEST FOR THE DESMOPLASTIC  
 ? NUMBER OF SEQUENCES: 21  
 ? CORRESPONDENCE ADDRESS:  
 ? ADDRESSEE: Cooper & Dunham LLP  
 ? STREET: 1185 Avenue of the Americas  
 ? CITY: New York  
 ? STATE: New York  
 ? COUNTRY: U.S.A.  
 ? ZIP: 10036  
 ? COMPUTER READABLE FORM:  
 ? MEDIUM TYPE: Floppy disk  
 ? COMPUTER: IBM PC compatible  
 ? OPERATING SYSTEM: PC-DOS/MS-DOS  
 ? SOFTWARE: PatentIn Release #1.0, Version #1.30  
 ? CURRENT APPLICATION DATA:  
 ? APPLICATION NUMBER: US/08/437,027  
 ? FILING DATE:  
 ? CLASSIFICATION: 536

OY	1691	GGCTTCCTCCCGCACCCTTTCGCGCCCGGGGGTGATCGAGGAGAGAGTGGCCCTGGT	1750
Db	33661	GACTGCTTTGGCCCTGCTATTTCTCACCTTAAGTGGGATCGTGGCAGAGTGGCCCTGGT	333220
OY	1751	GGCATGCGGGAGGAAAGAGTGGCCCTCATGATCGTGATGTCCTCGGTGAAATGTTCAGA	1810
Db	33321	GGCATGCGGGAGGAAAGAGTGGCCCTCATGATCGTGATGTCCTCGGTGAAATGTTCAGA	33380
OY	1811	GGTGGCCGTGTGGAACACAGGTGGCTTCCTGTGTGGCCGGGGCATGGAACCGAGTGGC	1870
Db	33381	GGTGGCCGTGTGGAACACAGAGTGGCTTCCTGTGTGGCCGGGGCATGGAACCGAGTGGC	33440
OY	1871	TTTGTGTGAGGAAGACGAGGTGGCCCTGGGGGGCCCTCGAGACTTTGATGGAACGATG	1930
Db	33441	TTTGTGTGAGGAAGACGAGGTGGCCCTGGGGGGCCCTCGAGACTTTGATGGAACGATG	33500
OY	1931	GGAGGAAGAAAGAGAGAGAGTGGAGACCTGGAAATATGATTA	1974
Db	33501	GGAGGAAGAAAGAGAGAGAGTGGAGACCTGGAAATATGATTAAGTGTGTGAA	33560
OY	1975	-----	1974
Db	33561	AGCAGCTGTGGCGCCGACAGGACAGTMAAGAGACACCCTTCCAGCTTGTGGCGCA	33620
OY	1975	-----	1974
Db	33621	GTCCTCATGTCCTTAGAAGCTTGTGATATGTGGTGTGGAGAGACGAGAGGGGACCTG	33680
OY	1975	-----	1974
Db	33681	GGGGCTCTGGAAGGGGCTTCTCACCCCTTCCATCTAACGAGGGGCCCTTTAAGCTT	33740
OY	1975	----AGCGGACACCGTCAGAGGCGGAGAGATGGGCCCTTAATGATGACAGAGCCCCGA	2030
Db	33741	GCAAGAGCGGACACCGTCAGAGGCGGAGAGATGGGCCCTTAATGATGACAGAGCCCCGA	33800
OY	2031	GAGCTGATGACATCAACAGATTAATTTTAAACGAAATGTTTAAATTAATTAATTC	2090
Db	33801	GAGCTGATGACATCAACAGATTAATTTTAAACGAAATGTTTAAATTAATTAATTC	33860



121 TCAAGATATGCA CAGACCA CCGAGGCA TATGGGCA CAAGA GCTATGGA ACTATGACA 180  
121 TCAAGATATGCA CAGACCA CCGAGGCA TATGGGCA CAAGA GCTATGGA ACTATGACA 180  
181 GCCCACTGATGTGAGCTATA CCGAGGCTAGA CCACTGCA ACTATGAGGAGAC CCGCTTA 240  
181 GCCCACTGATGTGAGCTATA CCGAGGCTAGA CCACTGCA ACTATGAGGAGAC CCGCTTA 240  
241 TGAACCTCTTATGAGACGCTCCCACTGGTTATACCTACCTCAACTGCCCCCAAGGATA 300  
241 TGAACCTCTTATGAGACGCTCCCACTGGTTATACCTACCTCAACTGCCCCCAAGGATA 300  
301 CAGCAGACCTGTC CAGGGGATATGGA CTA TGGTCTTATGATACCA CCACTGCTACAGTAC 360  
301 CAGCAGACCTGTC CAGGGGATATGGA CTA TGGTCTTATGATACCA CCACTGCTACAGTAC 360  
301 CAGCAGACCTGTC CAGGGGATATGGA CTA TGGTCTTATGATACCA CCACTGCTACAGTAC 360  
361 CACCA CCGAGGCTCTTATGAGCTAGTCTGCA TATGGA CTA TCA GCTGCTCTTATTCAGC 420  
361 CACCA CCGAGGCTCTTATGAGCTAGTCTGCA TATGGA CTA TCA GCTGCTCTTATTCAGC 420  
421 CTATGGGACAGCAGCAGCA CCACTGCA CCACTGCA CCGGAGATGGA AACAAGCCAC 480  
421 CTATGGGACAGCAGCAGCA CCACTGCA CCACTGCA CCGGAGATGGA AACAAGCCAC 480  
481 TGAAGACTAGTCAACCTCAATCTAGCA CAGAGGAGTTTACA CCAAGCCAGCTTACAGTATG 540  
481 TGAAGACTAGTCAACCTCAATCTAGCA CAGAGGAGTTTACA CCAAGCCAGCTTACAGTATG 540  
541 ACAAGATTAATCAAGTTATCCCAAGGTA CTTGGAGCTACCCCA TEGAGCCAGTCACTGC 600  
541 ACAAGATTAATCAAGTTATCCCAAGGTA CTTGGAGCTACCCCA TEGAGCCAGTCACTGC 600  
601 ACCTCCATCTCTACCTCTACAGCTATCTTCTCTACAGCA CCACTGATATATGATCAG 660  
601 ACCTCCATCTCTACCTCTACAGCTATCTTCTCTACAGCA CCACTGATATATGATCAG 660  
661 CAGTTACTCTCAGAGAA CACTATGGGCA CCGAGCACTATGGA CAGCA GAGTACTA 720  
661 CAGTTACTCTCAGAGAA CACTATGGGCA CCGAGCACTATGGA CAGCA GAGTACTA 720  
721 TGGTCAACAAAGCACTATGGGCA CCGAGCTCTCCCACTAGTTACC CCACTGATC 780  
721 TGGTCAACAAAGCACTATGGGCA CCGAGCTCTCCCACTAGTTACC CCACTGATC 780  
781 CTACAGCAAGCTCAAGTCAATATAGCA CAGAGAGCACTAGGAGGAGAGAGATTC 840  
781 CTACAGCAAGCTCAAGTCAATATAGCA CAGAGAGCACTAGGAGGAGAGAGATTC 840  
841 ATTCCGACAGGACCA CCGAGTATGAGTGTATTATGGGCA GAGATCTGAGGATTTTC 900  
841 ATTCCGACAGGACCA CCGAGTATGAGTGTATTATGGGCA GAGATCTGAGGATTTTC 900  
901 CCGACAGAGAGAA CCGAGATGAGTGTCTGATTA CCGGGGCA GGGGAGAGAGAGG 960  
901 CCGACAGAGAGAA CCGAGATGAGTGTCTGATTA CCGGGGCA GGGGAGAGAGAGG 960  
961 ATTGATCTGAGAGGCA TGAAGAGTGGGCGGGAGAGAGAGAGAGAGAGAGAGAGAG 1020  
961 ATTGATCTGAGAGGCA TGAAGAGTGGGCGGGAGAGAGAGAGAGAGAGAGAGAGAG 1020  
1021 CCGTGAAGAGAGAGTGGCTTCAATAAGCTGTGGA CCACTGATGAGAGAGAGAGAGAT 1080  
1021 CCGTGAAGAGAGAGTGGCTTCAATAAGCTGTGGA CCACTGATGAGAGAGAGAGAGAT 1080  
1081 TGAATGAGGCGCA CCGTATGATCAAGTGAAGCTGGA CCAAGTCAATTTATATGACA 1140  
1081 TGAATGAGGCGCA CCGTATGATCAAGTGAAGCTGGA CCAAGTCAATTTATATGACA 1140  
1141 AGAATTAATGAGAGTGTGACTATGATGATGATGATGATGATGATGATGATGATGATGAT 1200  
1141 AGAATTAATGAGAGTGTGACTATGATGATGATGATGATGATGATGATGATGATGATGAT 1200  
1201 TGTAAAGATGAACAAGAGAA CTGGGCAACCATGATCAATCTTACCTGGA CAGAGAAAC 1260

1201 TGTAAAGATGAACAAGAGAA CTGGGCAACCATGATCAATCTTACCTGGA CAGAGAAAC 1260  
1261 AGGAAAGCCCAAGAGGATGCCAGATGCTCTATGAAG CCACTGCACTGCAAGCTGTC 1320  
1261 AGGAAAGCCCAAGAGGATGCCAGATGCTCTATGAAG CCACTGCACTGCAAGCTGTC 1320  
1321 CGTGAAGTGGTTGATGGGAAAGATTTTCAAGGAGCAAACTTAAAGTCTCCCTGCTG 1380  
1321 CGTGAAGTGGTTGATGGGAAAGATTTTCAAGGAGCAAACTTAAAGTCTCCCTGCTG 1380  
1381 GAAAGAGCTCCCAATGAACAGTATGAGGAGTGTGCA CCGCTGAGAGGAGAGAGAT 1440  
1381 GAAAGAGCTCCCAATGAACAGTATGAGGAGTGTGCA CCGCTGAGAGGAGAGAGAT 1440  
1441 GCCACCA CCACTCTGAGAGAGTCCAGAGAGCCCAAGAGTCTCTGGGGAGACCA TGGGTG 1500  
1441 GCCACCA CCACTCTGAGAGAGTCCAGAGAGCCCAAGAGTCTCTGGGGAGACCA TGGGTG 1500  
1501 CATGGAGGCGGTGAGAGATGAGAGAGCTTCCCTCCAAAGAGAG CCGCGGGGTTCCCG 1560  
1501 CATGGAGGCGGTGAGAGATGAGAGAGCTTCCCTCCAAAGAGAG CCGCGGGGTTCCCG 1560  
1561 AGGAAAGCTCTGAGAGAGAA CAGTCA GCA CCGAGCTGAGAGCTGCGAGTGTCCAA 1620  
1561 AGGAAAGCTCTGAGAGAGAA CAGTCA GCA CCGAGCTGAGAGCTGCGAGTGTCCAA 1620  
1621 TCGGGTGTGGA AACAAGACTTCCCTGAGAGACAGAGTGAACA CAGTGTAAAGCTCC 1680  
1621 TCGGGTGTGGA AACAAGACTTCCCTGAGAGACAGAGTGAACA CAGTGTAAAGCTCC 1680  
1681 AAGAGCTGAAGGCTTCTCCCGCA CCGCTTCCGAGCCCGGAGTGTGATCTGAGAGAG 1740  
1681 AAGAGCTGAAGGCTTCTCCCGCA CCGCTTCCGAGCCCGGAGTGTGATCTGAGAGAG 1740  
1741 TGGCCTGTGAGTCCGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1800  
1741 TGGCCTGTGAGTCCGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1800  
1801 AATGTTCAAGAGTGGGCTGTGAGAGACAGAGTGTCTCCGTGTGAGCGGGGCAATGGA 1860  
1801 AATGTTCAAGAGTGGGCTGTGAGAGACAGAGTGTCTCCGTGTGAGCGGGGCAATGGA 1860  
1861 CCGAGGTGCTTGTGAT 1920  
1861 CCGAGGTGCTTGTGAT 1920  
1861 CCGAGGTGCTTGTGAT 1920  
1861 CCGAGGTGCTTGTGAT 1920  
1921 GGAACAGATGGAG 1980  
1921 GGAACAGATGGAG 1980  
1981 GCAAGGTGCTTGTGAT 2040  
1981 GCAAGGTGCTTGTGAT 2040  
2041 GACTACAGATTTATTTTAAACAGAAATGTTTAAATTTTAAATTTTAAATTTTAAATTTTAA 2100  
2041 GACTACAGATTTATTTTAAACAGAAATGTTTAAATTTTAAATTTTAAATTTTAAATTTTAA 2100  
2101 ATGTGGCCCA CCACTATGATTAATCTGTTCTGATCTTATGATTTTCA CCACTTGT 2160  
2101 ATGTGGCCCA CCACTATGATTAATCTGTTCTGATCTTATGATTTTCA CCACTTGT 2160  
2161 GAGAAACATTTAAACAGTTAAATGATGATGATGATGATGATGATGATGATGATGATGAT 2220  
2161 GAGAAACATTTAAACAGTTAAATGATGATGATGATGATGATGATGATGATGATGATGAT 2220  
2221 AAAAATGATTTTAAAGCTTTAAACAGTTAAAGCTTTGAGAGAGCTCAGATGAT 2280  
2221 AAAAATGATTTTAAAGCTTTAAACAGTTAAAGCTTTGAGAGAGCTCAGATGAT 2280  
2281 GTGAGAA CCAAGAGGCTCTTAACTGTAACAATGTTCAATGATGATGATGATGATGATGAT 2340



```

QY 1621 TCCGGGTTGTAAGAAACAGAACTTCGCTGAGAGAACAGATGCAACAGATGAAGCCCC 1680
DB 1621 TCCGGGTTGTAAGAAACAGAACTTCGCTGAGAGAACAGATGCAACAGATGAAGCCCC 1680
QY 1681 AAAGCCTGAAGGCTTCTCCGCAACCTTTCCGCCCCCGGGTGTGATGTCGCAAGG 1740
DB 1681 AAAGCCTGAAGGCTTCTCCGCAACCTTTCCGCCCCCGGGTGTGATGTCGCAAGG 1740
QY 1741 TGGCCCTGTGTGATGAGGAGGAGAGAGAGTGGCTCTCACTGATGTCGTCGCGGTG 1800
DB 1741 TGGCCCTGTGTGATGAGGAGGAGAGAGTGGCTCTCACTGATGTCGTCGCGGTG 1800
QY 1801 AAGTTTCAAGATGAGCGCTGTGAGAGACAGAGGTGCTTCGTCGTGCGCGGACATGA 1860
DB 1801 AAGTTTCAAGATGAGCGCTGTGAGAGACAGAGGTGCTTCGTCGTGCGCGGACATGA 1860
QY 1861 CCGAGGTGCTTTGTGTGAGAGAGACAGAGGTGCTTCGCGGAGGCCCCCTGACCTTTGAT 1920
DB 1861 CCGAGGTGCTTTGTGTGAGAGAGACAGAGGTGCTTCGCGGAGGCCCCCTGACCTTTGAT 1920
QY 1921 GGAACAGATGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1980
DB 1921 GGAACAGATGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1980
QY 1981 GCAACGTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2040
DB 1981 GCAACGTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2040
QY 2041 GACTACCAAGATTTATTTTAAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2100
DB 2041 GACTACCAAGATTTATTTTAAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2100
QY 2101 ATGTGGCCCAACATTAATATATATATATATATATATATATATATATATATATAT 2160
DB 2101 ATGTGGCCCAACATTAATATATATATATATATATATATATATATATATATATAT 2160
QY 2161 GAAGAAACATTAATAACAGTTAAATGTAGTGTGCGAGTTTCTTCTCTCTT 2220
DB 2161 GAAGAAACATTAATAACAGTTAAATGTAGTGTGCGAGTTTCTTCTCTCTT 2220
QY 2221 AAAAATGTTTGAAGACTTTAAACATGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2280
DB 2221 AAAAATGTTTGAAGACTTTAAACATGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2280
QY 2281 GTGAGAGACCAAGAGGCTCTTAACGTGAACATGTTCAATGTTGAGTGTGTTT 2340
DB 2281 GTGAGAGACCAAGAGGCTCTTAACGTGAACATGTTCAATGTTGAGTGTGTTT 2340
QY 2341 TTTTAAAAATTAATAATCCAAATGTTTAAATTAATAATAATAATAATAATAATA 2390
DB 2341 TTTTAAAAATTAATAATCCAAATGTTTAAATTAATAATAATAATAATAATAATA 2390

```

```

RESULT 3
US-09-960-706-1081
; Sequence 1081, Application US/09960706
; Publication No. US20030134280A1
; GENERAL INFORMATION:
; APPLICANT: Munger, William E;
; TITLE OF INVENTION: Identifying Drugs for and Diagnosis of Benign Prostatic Hyperplasia
; FILE REFERENCE: 44921-5029-01US
; CURRENT APPLICATION NUMBER: US/09/960,706
; CURRENT FILING DATE: 2001-09-24
; PRIOR APPLICATION NUMBER: 60/223,323
; PRIOR FILING DATE: 2000-08-07
; PRIOR APPLICATION NUMBER: 09/873,319
; PRIOR FILING DATE: 2001-06-05
; NUMBER OF SEQ ID NOS: 1124
; SOFTWARE: Patent Ver. 2.1
; SEQ ID NO 1081
; LENGTH: 2390
; TYPE: DNA

```

```

; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20030134280A1 X66899
US-09-960-706-1081

```

```

Query Match 99.9%; Score 2388.4; DB 10; Length 2390;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2389; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

QY 1 AGAGGAGACGAGACGTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 60
DB 1 AGAGGAGACGAGACGTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 60
QY 61 CAGTACCTTATGAGCAAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 120
DB 61 CAGTACCTTATGAGCAAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 120
QY 121 TCAGGATATGCAACAGACCAACCGAGCATATGGGCAACAAAGCTATGAACTTATGACA 180
DB 121 TCAGGATATGCAACAGACCAACCGAGCATATGGGCAACAAAGCTATGAACTTATGACA 180
QY 181 GCCCATGATGTACGATATATCCAGGCTCAGACCACTGCAACCTATGGGAGACCGCTTA 240
DB 181 GCCCATGATGTACGATATATCCAGGCTCAGACCACTGCAACCTATGGGAGACCGCTTA 240
QY 241 TGCATCTTATGAGACAGAGCTCCCACTGATATATCTCAACTGCCCCCAGGACATA 300
DB 241 TGCATCTTATGAGACAGAGCTCCCACTGATATATCTCAACTGCCCCCAGGACATA 300
QY 301 CAGCCAGCTGTTCAGAGGATATGCACTGTGTATATGATACCACTGTCTACAGTAC 360
DB 301 CAGCCAGCTGTTCAGAGGATATGCACTGTGTATATGATACCACTGTCTACAGTAC 360
QY 361 CAGCCAGCTGTTCAGAGGATATGCACTGTGTATATGATACCACTGTCTACAGTAC 420
DB 361 CAGCCAGCTGTTCAGAGGATATGCACTGTGTATATGATACCACTGTCTACAGTAC 420
QY 421 CTATGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 480
DB 421 CTATGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 480
QY 481 TGAAGTATGCACTCAATCTTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 540
DB 481 TGAAGTATGCACTCAATCTTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 540
QY 541 ACAGATTAACATGATATATCCAGGATACCTGAGAGAGAGAGAGAGAGAGAGAGAGAG 600
DB 541 ACAGATTAACATGATATATCCAGGATACCTGAGAGAGAGAGAGAGAGAGAGAGAGAG 600
QY 601 ACCTCATCTACCTCTCTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 660
DB 601 ACCTCATCTACCTCTCTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 660
QY 661 CAGTTAATCTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 720
DB 661 CAGTTAATCTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 720
QY 721 TGGTCAACAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 780
DB 721 TGGTCAACAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 780
QY 781 CTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 840
DB 781 CTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 840
QY 841 ATTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 900
DB 841 ATTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 900
QY 901 CGAGACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 960
DB 901 CGAGACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 960

```





Db 241 TGCACTTCTTATGAGACGCTCCACTGGTTATTAATCTCAACTGCCCCCGAGGATA 300  
Qy 301 CAGCCAGCCTGTCCAGGGGATATGCACTGGTCTTATGATACCACTGCTCACTGAC 360  
Db 301 CAGCCAGCCTGTCCAGGGGATATGCACTGGTCTTATGATACCACTGCTCACTGAC 360  
Qy 361 CACCAACCCAGGCTCTCTATGAGCTCACTGATATGAGCACTGAGCTCTTATGACAG 420  
Db 361 CACCAACCCAGGCTCTCTATGAGCTCACTGATATGAGCACTGAGCTCTTATGACAG 420  
Qy 421 CTATGGGACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 480  
Db 421 CTATGGGACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 480  
Qy 481 TGAGACTATGCTCACTCAATCTAGCAGAGGGGGTTTAAACAGCCGCTTAGATATGG 540  
Db 481 TGAGACTATGCTCACTCAATCTAGCAGAGGGGGTTTAAACAGCCGCTTAGATATGG 540  
Qy 541 ACAAGATTAATCACTATGCTCCAGGTACCTGGAGCTACCCCACTGAGCAGTCACTGC 600  
Db 541 ACAAGATTAATCACTATGCTCCAGGTACCTGGAGCTACCCCACTGAGCAGTCACTGC 600  
Qy 601 ACCCTCATCTACCTCTCTACAGCTATTCCTCTACACAGCCGACTAGTATGATCAGAG 660  
Db 601 ACCCTCATCTACCTCTCTACAGCTATTCCTCTACACAGCCGACTAGTATGATCAGAG 660  
Qy 661 CAGTTACTCTCAGAGAAACACTATGGGCAACCGAGAGCTATGGACAGAGTATGCTA 720  
Db 661 CAGTTACTCTCAGAGAAACACTATGGGCAACCGAGAGCTATGGACAGAGTATGCTA 720  
Qy 721 TGGTCAACAAAGCAGCTATGGGAGCAGCCCTCCCACTAGTTACCCCACTGATC 780  
Db 721 TGGTCAACAAAGCAGCTATGGGAGCAGCCCTCCCACTAGTTACCCCACTGATC 780  
Qy 781 CTACAGCCAAAGCTCTCAAGTCAATATAGCAACAGAGCAGAGCTACCGGAGCAGAGTTC 840  
Db 781 CTACAGCCAAAGCTCTCAAGTCAATATAGCAACAGAGCAGAGCTACCGGAGCAGAGTTC 840  
Qy 841 ATTTCGACAGAGCCAGCTCAGTATGAGTGGTGTATATGGGCAAGAGTCTGAGAGATTTTC 900  
Db 841 ATTTCGACAGAGCCAGCTCAGTATGAGTGGTGTATATGGGCAAGAGTCTGAGAGATTTTC 900  
Qy 901 CGGACAGAGAGAAACCGGAGCATGATGGCCCTGATTAACCGGGGAGGGGAGAGGGGG 960  
Db 901 CGGACAGAGAGAAACCGGAGCATGATGGCCCTGATTAACCGGGGAGGGGAGAGGGGG 960  
Qy 961 ATTGATCTGTGAGAGCATGAGCAGAGGTGGCGGGAGAGAGAGCGGCTGGAATGGGAG 1020  
Db 961 ATTGATCTGTGAGAGCATGAGCAGAGGTGGCGGGAGAGAGAGCGGCTGGAATGGGAG 1020  
Qy 1021 CGCTGAGAGAGAGGTGGCTTCAATTAAGCTGTGGACCCATGGATGAAGAGCCAGATCT 1080  
Db 1021 CGCTGAGAGAGAGGTGGCTTCAATTAAGCTGTGGACCCATGGATGAAGAGCCAGATCT 1080  
Qy 1081 TGATCTAGGCCCACCTGTAGATCCAGATGAAGAATCTGACAAAGTGCATTTATATGACA 1140  
Db 1081 TGATCTAGGCCCACCTGTAGATCCAGATGAAGAATCTGACAAAGTGCATTTATATGACA 1140  
Qy 1141 AGGATTAATGACAGTGTGACTAGATCTGAGAGCTTCTTTAAGCAGTGGGGT 1200  
Db 1141 AGGATTAATGACAGTGTGACTAGATCTGAGAGCTTCTTTAAGCAGTGGGGT 1200  
Qy 1201 TGTTAAGATGAACAGAGAACTGGGCAACCAATGATCCATCTGAGCAAGAGAAAC 1260  
Db 1201 TGTTAAGATGAACAGAGAACTGGGCAACCAATGATCCATCTGAGCAAGAGAAAC 1260  
Qy 1261 AGGAAAGCCCAAGAGGAGTGCAGAGTCTCTATGAAGAACCCCACTGCAAGGCTGC 1320  
Db 1261 AGGAAAGCCCAAGAGGAGTGCAGAGTCTCTATGAAGAACCCCACTGCAAGGCTGC 1320  
Qy 1321 CGTGAATGTGTTTATGAGGAAAGATTTTCAAGGAGCAAACTTAAAGTCTCCCTGCTGC 1380  
Db 1321 CGTGAATGTGTTTATGAGGAAAGATTTTCAAGGAGCAAACTTAAAGTCTCCCTGCTGC 1380

Qy 1381 GAAGAAAGCTCCAAATGAACAGTATGGGGGTGTCTGACACCCCGTGAAGGAGCAGAGCAT 1440  
Db 1381 GAAGAAAGCTCCAAATGAACAGTATGGGGGTGTCTGACACCCCGTGAAGGAGCAGAGCAT 1440  
Qy 1441 GCCACCAACACTCCGTGAGAGTCCAGAGAGCCCAAGAGGTCTGTGGGAGACCATGGGTGC 1500  
Db 1441 GCCACCAACACTCCGTGAGAGTCCAGAGAGCCCAAGAGGTCTGTGGGAGACCATGGGTGC 1500  
Qy 1501 CATGGAGAGCCGTGAGAGAGATAGAGAGAGCTTCCCTCCAAAGAGACCCCGGGGTCCCG 1560  
Db 1501 CATGGAGAGCCGTGAGAGAGATAGAGAGAGCTTCCCTCCAAAGAGAGCCCGGGGTCCCG 1560  
Qy 1561 AGGGAAGCCCTCTGAGAGAGAAAGTCCAGACCCAGAGCTGAGAGCTGAGAGTGTCCAA 1620  
Db 1561 AGGGAAGCCCTCTGAGAGAGAAAGTCCAGACCCAGAGCTGAGAGCTGAGAGTGTCCAA 1620  
Qy 1621 TCCGGGTGTGAAACAGAACTTCGCTGAGAAACAGAGTGCACAGTGAAGCCCC 1680  
Db 1621 TCCGGGTGTGAAACAGAACTTCGCTGAGAAACAGAGTGCACAGTGAAGCCCC 1680  
Qy 1681 AAAGCTGAAGGCTTCTCCCGCACCTTTCCGCCCCGGGTGTGATCTGTGGCAGAGG 1740  
Db 1681 AAAGCTGAAGGCTTCTCCCGCACCTTTCCGCCCCGGGTGTGATCTGTGGCAGAGG 1740  
Qy 1741 TGGCCTGTGGGATGCGGGGAGAGAGAGTGGCTCATGATCTGTGTGCTCCGGTGG 1800  
Db 1741 TGGCCTGTGGGATGCGGGGAGAGAGAGTGGCTCATGATCTGTGTGCTCCGGTGG 1800  
Qy 1801 AATGTTCAAGAGTGGCCGTGTGTGAGACAGAGTGGCTTCCGTGTGTGGCGGGCAATGA 1860  
Db 1801 AATGTTCAAGAGTGGCCGTGTGTGAGACAGAGTGGCTTCCGTGTGTGGCGGGCAATGA 1860  
Qy 1861 CCGAGGTGGCTTGTGTGTGAGAGAGAGAGTGGCTTGGGGGGCCCCCTGAGACTTGTAT 1920  
Db 1861 CCGAGGTGGCTTGTGTGTGAGAGAGAGAGTGGCTTGGGGGGCCCCCTGAGACTTGTAT 1920  
Qy 1921 GGAACAGATGGGAGAAAG 1980  
Db 1921 GGAACAGATGGGAGAAAG 1980  
Qy 1981 GCAACGTCAGAGAGCGAGAGATGGGCTTACTAGATGAGAGAGCCCGCAGAGCTGACT 2040  
Db 1981 GCAACGTCAGAGAGCGAGAGATGGGCTTACTAGATGAGAGAGCCCGCAGAGCTGACT 2040  
Qy 2041 GACTACCAAGTTTATTTTAAACAGAAATGTTTAAATTTTAAATTTTCAATTTATA 2100  
Db 2041 GACTACCAAGTTTATTTTAAACAGAAATGTTTAAATTTTAAATTTTCAATTTATA 2100  
Qy 2101 ATGTGGCACAACATTAATGATTAATCTTGTCTGTACTTATGATTTTCAACATTTGT 2160  
Db 2101 ATGTGGCACAACATTAATGATTAATCTTGTCTGTACTTATGATTTTCAACATTTGT 2160  
Qy 2161 GAAGAAACATTTAAACAGTTAATGATGATGTGGAGTTTCTTCTCTCTCTCT 2220  
Db 2161 GAAGAAACATTTAAACAGTTAATGATGATGTGGAGTTTCTTCTCTCTCTCTCT 2220  
Qy 2221 AAAAATGTTGTTTAAAGCTTTTAAACAATGGGAAACCCCTGTGAGCATGCTCAATGAT 2280  
Db 2221 AAAAATGTTGTTTAAAGCTTTTAAACAATGGGAAACCCCTGTGAGCATGCTCAATGAT 2280  
Qy 2281 GTGAGAGAACAGAGGCGCTTAACTGTAAACAATGCTTCAATGATGATGATTTTCTT 2340  
Db 2281 GTGAGAGAACAGAGGCGCTTAACTGTAAACAATGCTTCAATGATGATGATTTTCTT 2340  
Qy 2341 TTTTAAATTAATTAATTTCCAAATGTTTAAATTAATTAATTAATTAATTAATTAATTA 2390  
Db 2341 TTTTAAATTAATTAATTTCCAAATGTTTAAATTAATTAATTAATTAATTAATTAATTA 2390

RESULT 5  
US-09-822-830A-410/c  
Sequence 410, Application US/0982830A

Patent No. US20020142952A1  
 GENERAL INFORMATION:  
 APPLICANT: Genetics Institute, Inc.  
 APPLICANT: Wong, Gordon G.  
 APPLICANT: Clark, Hilary  
 APPLICANT: Fachel, Kim  
 APPLICANT: Agostino, Michael J.  
 APPLICANT: Hower, Steven H.  
 APPLICANT: Resnick, Richard J.  
 APPLICANT: Gulukota, Kamalakkar  
 APPLICANT: Graham, James R.  
 TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING NOVEL SECRETED PROTEINS  
 FILE REFERENCE: GIN 6402  
 CURRENT APPLICATION NUMBER: US/09/822,830A  
 CURRENT FILING DATE: 2001-03-29  
 PRIOR APPLICATION NUMBER: 60/195,604  
 PRIOR FILING DATE: 2000-04-06  
 NUMBER OF SEQ ID NOS: 631  
 SOFTWARE: PatentIn Ver. 2.0  
 SEQ ID NO 410  
 LENGTH: 2273  
 TYPE: DNA  
 ORGANISM: Homo sapiens  
 US-09-822-830A-410

Query Match 91.2%; Score 2179.6; DB 9; Length 2273;  
 Best Local Similarity 97.3%; Pred. No. 0;  
 Matches 2259; Conservative 0; Mismatches 9; Indels 54; Gaps 2;

42 AAATGGGCTCAGGATTACCTATGACCAAGCTGAGGCGACAGGGCTACAGT 101  
 2268 AAGGTGCTCAGGATTACCTATGACCAAGCTGAGGCGACAGGGCTACAGT 2209  
 102 CTACACCGCCGACCCACTCAAGATATGACAGACCAACCAGGCATATGGCAAAA 161  
 2208 CTACACCGCCGACCCACTCAAGATATGACAGACCAACCAGGCATATGGCAAAA 2149  
 162 GCTATGGAACCTATGACAGCCCACTGATGTCACTATACCCAGGCTCAACCACTGCA 221  
 2148 GCTATGGAACCTATGACAGCCCACTGATGTCACTATACCCAGGCTCAACCACTGCA 2089  
 222 CCTATGGGAGACCGGCTATGACCTTATGAGAGCGCTCCCACTGTTATATCTC 281  
 2088 CCTATGGGAGACCGGCTATGACCTTATGAGAGCGCTCCCACTGTTATATCTC 2029  
 282 CAATGCCCCCAGGACATAGAGCGCTGTCAGAGGGTATGGCACTGCTTATGATA 341  
 2028 CAATGCCCCCAGGACATAGAGCGCTGTCAGAGGGTATGGCACTGCTTATGATA 1969  
 342 CCACCACTGCTACAGTACCAACCAAGGCTCTCTATGACGCTCACTTGCATATGGCA 401  
 1968 CCACCACTGCTACAGTACCAACCAAGGCTCTCTATGACGCTCACTTGCATATGGCA 1909  
 402 CTGAGCTGCTTATCCAGCTATGGGAGAGCGACAGGCACTGACCTTCAAGACCGC 461  
 1908 CTGAGCTGCTTATCCAGCTATGGGAGAGCGACAGGCACTGACCTTCAAGACCGC 1849  
 462 AGGATGGAACAAGCCCACTGAGACTAGTCACTCACTCACTTACGACAGGGGTTACAAC 521  
 1848 AGGATGGAACAAGCCCACTGAGACTAGTCACTCACTCACTTACGACAGGGGTTACAAC 1789  
 522 AGCCCACTGAGATATGACAGAGTATGACAGTATATCCCAAGTATCTTGGAGTATCC 581  
 1788 AGCCCACTGAGATATGACAGAGTATGACAGTATATCCCAAGTATCTTGGAGTATCC 1729  
 582 CGATGAGCAGATGACCTTCACTCACTCACTCACTCACTCACTCACTCACTCACTCACT 641  
 1728 CGATGAGCAGATGACCTTCACTCACTCACTCACTCACTCACTCACTCACTCACTCACT 1669  
 642 CGATGAGCAGATGACCTTCACTCACTCACTCACTCACTCACTCACTCACTCACTCACT 701  
 1668 CGATGAGCAGATGACCTTCACTCACTCACTCACTCACTCACTCACTCACTCACTCACT 1609

702 ATGACAGCAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 761  
 1608 ATGACAGCAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1549  
 762 ACCACCCCAACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 821  
 1548 ACCACCCCAACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1489  
 822 GCTACGGGAGCAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 881  
 1488 GCTACGGGAGCAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1429  
 882 AGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 941  
 1428 AGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1369  
 942 GGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1001  
 1368 GGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1309  
 1002 GACGCGGTGGAATGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1061  
 1308 GACGCGGTGGAATGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1252  
 1062 TGATGAGAGCAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1121  
 1251 TGATGAGAGCAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1192  
 1122 ACAGTGAATTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1181  
 1191 ACAGTGAATTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1132  
 1182 TCTTAAAGCAGTGGGAGTGTAAAGTGAAGAAAGAAAGTGGGCAACCCATGATGACA 1241  
 1131 TCTTAAAGCAGTGGGAGTGTAAAGTGAAGAAAGAAAGTGGGCAACCCATGATGACA 1072  
 1242 TCTTAAAGCAGTGGGAGTGTAAAGTGAAGAAAGTGGGCAACCCATGATGACA 1301  
 1071 TCTTAAAGCAGTGGGAGTGTAAAGTGAAGAAAGTGGGCAACCCATGATGACA 1012  
 1302 CACCACTGACCAAGGCTGCGGTGAATGATGATGATGATGATGATGATGATGATGATGAT 1361  
 1011 CACCACTGACCAAGGCTGCGGTGAATGATGATGATGATGATGATGATGATGATGATGAT 952  
 1362 TTAAGTCTCTGCTGCTGAGAGAGCTTCAATGAACAGTATGCGGGAGTGTGTCAC 1421  
 951 TTAAGTCTCTGCTGCTGAGAGAGCTTCAATGAACAGTATGCGGGAGTGTGTCAC 892  
 1422 CCGGTGAGGAGCAGAGGATGACCACTCCGTGAGAGTCCAGAGAGGCGCCAGAGAGTC 1481  
 891 CCGGTGAGGAGCAGAGGATGACCACTCCGTGAGAGTCCAGAGAGGCGCCAGAGAGTC 832  
 1482 CTGGGGGAGCAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1541  
 831 CTGGGGGAGCAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 772  
 1542 GAGGAGCCCGGGGTTCCGAGGAGACCCCTTCTGAGAGGAGAAAGTTCAGACACGAGCTG 1601  
 771 GAGGAGCCCGGGGTTCCGAGGAGACCCCTTCTGAGAGGAGAAAGTTCAGACACGAGCTG 712  
 1602 GAGACTGAGAGTCTCCCAATCCGGGTTGAGAGAAACAGAACTTGGCTGAGAGAAAGAGT 1661  
 711 GAGACTGAGAGTCTCCCAATCCGGGTTGAGAGAAACAGAACTTGGCTGAGAGAAAGAGT 652  
 1662 GCAACCACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1721  
 651 GCAACCACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 643  
 1722 GTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1781  
 642 GTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 583  
 1782 ATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1841

Db	Sequence	Score	DB	Length	Matches
Db	582 ATCGTGTGGTCCCGGTGGAAATGTTTCAGAGGTGGCCGTGTGGAGACAGAGTGGCTTCC	523			
Qy	1842 GTGGTGGCCCGGGGATGAGACCGAGGTGGCTTTGGTGGAGAAAGAGTGGCCCTTGGG	1901			
Db	522 GTGGTGGCCCGGGGATGAGACCGAGGTGGCTTTGGTGGAGAAAGAGTGGCCCTTGGG	463			
Qy	1902 GGCCCCCTTGACCTTTTGATGGAAACAGATGGAGAGAAAGAGAGAGACGTGGAGAGACTG	1961			
Db	462 GGCCCCCTTGACCTTTTGATGGAAACAGATGGAGAGAAAGAGAGAGACGTGGAGAGACTG	403			
Qy	1962 GAAAAATGATTAAGAGCGAGCAGCCGTCCAGAGCCGAGATCGGCCCTTACTAGATGAGAGA	2021			
Db	402 GAAAAATGATTAAGAGCGAGCAGCCGTCCAGAGCCGAGATCGGCCCTTACTAGATGAGAGA	343			
Qy	2022 GACCCCGAGAGCTGATGACTACAGATTTATTTTAAACAGAAAAATGTTTTAAAT	2081			
Db	342 GACCCCGAGAGCTGATGACTACAGATTTATTTTAAACAGAAAAATGTTTTAAAT	283			
Qy	2082 TTATATATTCATTTATTAATGTTGGCCCAACAATATATTTCTGTGCTGTACTTT	2141			
Db	282 TTATATATTCATTTATTAATGTTGGCCCAACAATATATTTCTGTGCTGTACTTT	223			
Qy	2142 AGTATTTTTCACATTTGTGTAAGAAACATTTAAACAAGTTAATGTTGTGGGAGATT	2201			
Db	222 AGTATTTTTCACATTTGTGTAAGAAACATTTAAACAAGTTAATGTTGTGGGAGATT	163			
Qy	2202 TTTTTCCTTCCTCTTTTAAAAATGTTGTTTAAAGACTTTAAGATGGAAACCCCTTGT	2261			
Db	162 TTTTTCCTTCCTCTTTTAAAAATGTTGTTTAAAGACTTTAAGATGGAAACCCCTTGT	103			
Qy	2262 GAGCATGCTCAGTATCATTTGTGAGAACCAAGAGGCGCTCTTAACTGTAAACAAATGTTTAT	2321			
Db	102 GAGCATGCTCAGTATCATTTGTGAGAACCAAGAGGCGCTCTTAACTGTAAACAAATGTTTAT	43			
Qy	2322 GGTGTGATGTTTTTTTTTTTTTTTAAATTAATTCGAA 2363				
Db	42 GGTGTGATGTTTTTTTTTTTTTTTAAATTAATTCGAA 1				

Db	2171	AAAGTGAAGAAATGGGGTCCAGGGAATTACGTA	CTTAATAGCCAAAGCTGACGGCCAGACGG	2112
QY	93	GCTACAGTGGTTACACCGCCCAAGCCCACTGA	AGAAATATGACAGACCACTCCAGGCATATG	152
Db	2111	GCTACAGTGGTTACACCGCCCAAGCCCACTGA	AGAAATATGACAGACCACTCCAGGCATATG	2052
QY	153	GGCAACAAAGCTATGGAACCTATGGAACGGCCA	CTGATGTCACCTATACCCAGGCTCAGA	212
Db	2051	GGCAACAAAGCTATGGAACCTATGGAACGGCCA	CTGATGTCACCTATACCCAGGCTCAGA	1992
QY	213	CCACTGCAACCTATGAGCAGACCGCCTATGCAA	CTTTCTATATGACAGCCCTCCACTGGTT	272
Db	1991	CCACTGCAACCTATGAGCAGACCGCCTATGCAA	CTTTCTATATGACAGCCCTCCACTGGTT	1932
QY	273	ATACTACTCCAACTGCCCCCAAGCATACAGCCAG	CTGTCCAGGGGTATGSCACTGGGTG	332
Db	1931	ATACTACTCCAACTGCCCCCAAGCATACAGCCAG	CTGTCCAGGGGTATGSCACTGGGTG	1872
QY	333	CTTAATGATACCAACACAGCTGCTACAGTACACA	CCCGAGGCGTCTCTATGAGAGCTCAAGCTG	392
Db	1871	CTTAATGATACCAACCTGCTACAGTACACAACCC	AGGCGTCTCTATGAGAGCTCAAGCTG	1812
QY	393	CATATGCGCACTAGCCTGCTTATTCAGCCTATGG	GGCGACGACCGCACTTGCACTTA	452
Db	1811	CATATGCGCACTAGCCTGCTTATTCAGCCTATGG	GGCGACGACCGCACTTGCACTTGCACTTA	1752
QY	453	CAAGACCGCAGAGATGGAACAAGCCCACTGACAT	GACTCACTCAATCTTAGACAGGGG	512
Db	1751	CAAGACCGCA - GATGGAACAAGCCCACTGACAT	GACTCACTCAATCTTAGACAGGGG	1693
QY	513	GTTTCAACACAGCCCGCTTGGATATGGAAGAAGT	AACTTACAGTTATCCCGAGGTACTG	572
Db	1692	GTTTCAACACAGCCCGCTTGGATATGGAAGAAGT	AACTTACAGTTATCCCGAGGTACTG	1633
QY	573	GGAAGCTACCCCATGACAGCAGTCACTGCACTCC	ATCATCTCACTCCCTTACCAAGCTATTTCT	632
Db	1632	GGAAGCTACCCCATGACAGCAGTCACTGCACTCC	ATCATCTCACTCCCTTACCAAGCTATTTCT	1573
QY	633	CTACACAGCCGACTAGTTATGATCAAGACAGTTA	CTCTCAGACAGAACCTATGGGCAAC	692
Db	1572	CTACACAGCCGACTAGTTATGATCAAGACAGTTA	CTCTCAGACAGAACCTATGGGCAAC	1513
QY	693	CGAGCAGCTATATGACACACAGATATGCTATGCT	CAACAACACAGTATGGGACAGAGCTC	752
Db	1512	CGAGCAGCTATATGACACACAGATATGCTATGCT	CAACAACACAGTATGGGACAGAGCTC	1453
QY	753	CCAATAAGTTAACCAACCCCAATCTGATCTCTA	CAGCCCAAGCTCCAAATGTAATAGCCAAC	812
Db	1452	CCAATAAGTTAACCAACCCCAATCTGATCTCTA	CAGCCCAAGCTCCAAATGTAATAGCCAAC	1393
QY	813	AGAGCAGAGCTACGGGCAAGCAGAGATTCAATT	CCGACAGACCAACCCAGTAGCTGGGTG	872
Db	1392	AGAGCAGAGCTACGGGCAAGCAGAGATTCAATT	CCGACAGACCAACCCAGTAGCTGGGTG	1333
QY	873	TTTATATGGCAGAGATCTGAGAGATTTTCCCGA	CCAGAGAGAAACCCGAGCATGATGGCC	932
Db	1332	TTTATATGGCAGAGATCTGAGAGATTTTCCCGA	CCAGAGAGAAACCCGAGCATGATGGCC	1273
QY	933	CTGATTAACCGGGGAGGGGGAATTTGATCTGTAG	GAGCAGATAGCAGAGGTGGCC	992
Db	1272	CTGATTAACCGGGGAGGGGGAATTTGATCTGTAG	GAGCAGATAGCAGAGGTGGCC	1213
QY	993	GGGGAGAGAGACGGGTGGAATGGGACAGCGCT	GTGAGACGAGAGCTTCATTAAGCTTG	1052
Db	1212	GGGGAGAGAGACGGGTGGAATGGGACAGCGCT	GTGAGAGAGAGCTTCATTAAGCTTG	1153
QY	1053	GTGAGACCATGGAATGAAGAACCAATCTTGATCT	TATAGGCCCACTGTGATTCACATGAAG	1112
Db	1152	GTGAGACCATGGAATGAAGAACCAATCTTGATCT	TATAGGCCCACTGTGATTCACATGAAG	10933
QY	1113	ACTCTGACAAAGTGCAATTTATGTAACAAGATTA	ATAATGACAGGTGACTCTAGATGATC	11727

Db 1092 ACTCTGACAAAGTGCATTTATGTACAGGATTAATGACAGTGTACTATGATATC 1033  
 Qy 1173 TGGCAGCTCTTTAAGCAGTGTGGGTGTGTAAAGATGAACAAGAACTGGGGAACCA 1232  
 Db 1032 TGGCAGCTCTTTAAGCAGTGTGGGTGTGTAAAGATGAACAAGAACTGGGGAACCA 973  
 Qy 1233 TGATCCACATCTACCTGGAAGAAAGAAAGCCAAAGGCGATGCCAGTGTCT 1292  
 Db 972 TGATCCACATCTACCTGGAAGAAAGAAAGCCAAAGGCGATGCCAGTGTCT 913  
 Qy 1293 ATGAAAGCCACCACTGCAAGGCTCCGTGGAATGATGGGAAAGATTTTCAAG 1352  
 Db 912 ATGAAAGCCACCACTGCAAGGCTCCGTGGAATGATGGGAAAGATTTTCAAG 853  
 Qy 1353 GAGCAAACTTAAAGTCTCTCTGTGTGGAAGAGCTCCAAATGAAGATGTGGGTG 1412  
 Db 852 GAGCAAACTTAAAGTCTCTCTGTGTGGAAGAGCTCCAAATGAAGATGTGGGTG 793  
 Qy 1413 GTCTGCCACCCCTGAGGGGAGAGGCAATGCAACCACTCCGTGAGGTCCAGGAGGC 1472  
 Db 792 GTCTGCCACCCCTGAGGGGAGAGGCAATGCAACCACTCCGTGAGGTCCAGGAGGC 733  
 Qy 1473 CAGAGGTCTCTGGGGAGCCCATGTGTCATGGAGGCGGTGAGAGATGAGAGGCT 1512  
 Db 732 CAGAGGTCTCTGGGGAGCCCATGTGTCATGGAGGCGGTGAGAGATGAGAGGCT 673  
 Qy 1533 TCCCTCCAAAGAACCCCGGGGTCTCCGAGGAAACCCCTCTGAGGAGAAAGCTCCAGC 1592  
 Db 672 TCCCTCCAAAGAACCCCGGGGTCTCCGAGGAAACCCCTCTGAGGAGAAAGCTCCAGC 613  
 Qy 1593 ACCGAGCTGAGAGCTGAGCAGTGTCCAAATCCGGGTGTGGAACCAAGCTTCCGAG 1652  
 Db 612 ACCGAGCTGAGAGCTGAGCAGTGTCCAAATCCGGGTGTGGAACCAAGCTTCCGAG 553  
 Qy 1653 GAAAGAGTGCACCAAGTGTAAAGCCCAAGCTTGAAGGCTTCTCCGCGCAACCTTTTC 1712  
 Db 552 GAAAGAGTGCACCAAGTGTAAAGCCCAAGCTTGAAGGCTTCTCCGCGCAACCTTTTC 493  
 Qy 1713 CGGCCCCGGGTGTGTGTGTGAGAGAGTGGCCCTGTGTGATGCGGGGAGAAAGAGTG 1772  
 Db 492 CGGCCCCGGGTGTGTGTGTGAGAGAGTGGCCCTGTGTGATGCGGGGAGAAAGAGTG 433  
 Qy 1773 GGCCTAGATGT 1832  
 Db 432 GGCCTAGATGT 373  
 Qy 1833 GTGGCTTCCGT 1892  
 Db 372 GTGGCTTCCGT 313  
 Qy 1893 GGCCTGGGGGGCCCTGTGACCTTTGTATGAAACAGATGGAGAAAGAAAGAGAGAGTG 1952  
 Db 312 GGCCTGGGGGGCCCTGTGACCTTTGTATGAAACAGATGGAGAAAGAAAGAGAGAGTG 253  
 Qy 1953 GAGGACCTGGAAGAAATGATGAAGGAGCAACCTGTGAGAGGAGAGAGAGAGAGAGTG 2012  
 Db 252 GAGGACCTGGAAGAAATGATGAAGGAGCAACCTGTGAGAGGAGAGAGAGAGAGAGTG 193  
 Qy 2013 AGATGAGTG 2072  
 Db 192 AGATGAGTG 133  
 Qy 2073 GTTTTAAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 2132  
 Db 132 GTTTTAAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 73  
 Qy 2133 CTGTACTTATGATTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 2189  
 Db 72 CTGTACTTATGATTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 16

RESULT 7  
 US-10-094-749-1074

; Sequence 1074, Application US/10094749  
 ; Publication No. US20030219741A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: ISOGAI, TAKAO  
 ; APPLICANT: SUGIYAMA, TOMOYASU  
 ; APPLICANT: OTSUKI, TETSUJI  
 ; APPLICANT: MAKAMATSU, AI  
 ; APPLICANT: SATO, HIROYUKI  
 ; APPLICANT: ISHII, SHIZUKO  
 ; APPLICANT: YAMAMOTO, JUN-ICHI  
 ; APPLICANT: ISONO, YUUKO  
 ; APPLICANT: HIO, YURI  
 ; APPLICANT: OTSUKA, KAORU  
 ; APPLICANT: NAGAI, KEIICHI  
 ; APPLICANT: IRIE, RYOTARO  
 ; APPLICANT: TAMECHIKA, ICHIRO  
 ; APPLICANT: SERI, NAOHICO  
 ; APPLICANT: YOSHIKAWA, TSUTOMU  
 ; APPLICANT: OTSUKA, MOTOKUKI  
 ; APPLICANT: NAGAHARI, KENJI  
 ; APPLICANT: MASUHO, YASUHIKO  
 ; TITLE OF INVENTION: NOVEL FULL-LENGTH cDNA  
 ; FILE REFERENCE: 084335/0160  
 ; CURRENT APPLICATION NUMBER: US/10/094,749  
 ; CURRENT FILING DATE: 2002-03-12  
 ; PRIOR APPLICATION NUMBER: 60/350,435  
 ; PRIOR FILING DATE: 2002-01-24  
 ; PRIOR APPLICATION NUMBER: JP 2001-328381  
 ; PRIOR FILING DATE: 2001-09-14  
 ; NUMBER OF SEQ ID NOS: 3381  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO 1074  
 ; LENGTH: 1988  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 US-10-094-749-1074  
 Query Match 75.7% Score 1808.4; DB 17; Length 1988;  
 Best Local Similarity 92.2% Pred. No. 0;  
 Matches 1987; Conservative 0; Mismatches 1; Indels 168; Gaps 1;  
 Db 33 AAGAGAGAAATGAGGCTCCAGGATTAAGTACTTATGACCAAGCTGAGGAGGAGG 92  
 1 AAGAGAGAAATGAGGCTCCAGGATTAAGTACTTATGACCAAGCTGAGGAGGAGG 60  
 Qy 93 GCTACAGTGTCTTACACCGCCAGCCCACTCAAGATATGACAGACCAAGGAGATATG 152  
 Db 61 GCTACAGTGTCTTACACCGCCAGCCCACTCAAGATATGACAGACCAAGGAGATATG 120  
 Qy 153 GGCACAAAGCTATGGAACCTATGAGACGCCACTGATGTCAAGTATACCAAGGCTCAGA 212  
 Db 121 GGCACAAAGCTATGGAACCTATGAGACGCCACTGATGTCAAGTATACCAAGGCTCAGA 180  
 Qy 213 CCACTGCAACTATGAGGAGACCGGCTATGCACTTATGAGACGGCTCCACTGGTT 272  
 Db 181 CCACTGCAACTATGAGGAGACCGGCTATGCACTTATGAGACGGCTCCACTGGTT 240  
 Qy 273 ATACTACTCAACTGCCCCCAGGAGATACAGGAGCTGTTCAGGGGTATGAGCACTGGTG 332  
 Db 241 ATACTACTCAACTGCCCCCAGGAGATACAGGAGCTGTTCAGGGGTATGAGCACTGGTG 300  
 Qy 333 CTATGATACCAACTGCTATGATGATGATGATGATGATGATGATGATGATGATGATGATG 392  
 Db 301 CTATGATACCAACTGCTATGATGATGATGATGATGATGATGATGATGATGATGATGATG 360  
 Qy 393 CATATGCACTGAGCGCTTATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 452  
 Db 361 CATATGCACTGAGCGCTTATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 420  
 Qy 453 CAAGACGAGAGATGGAACCAAGCCACTGAGAGTACTGAACTCAATCTTATGAGCAAGGGG 512  
 Db 421 CA----- 422



QY 513 GTTACAACAGCCAGCCTGATATGACAGATTAATACTAGTTATCCCAAGTACTG 572  
Db 423 ----- 422  
QY 573 GGAGCTACCCATGACAGCCAGTCACTGACCTCCATCTCACTCTCAACGATATCT 632  
Db 423 ----- AGCTATCTCT 432  
QY 633 CTACACAGCCGACTAGTTATGATCAAGAGATTAATCTCTCAAGCAAGAACCTATGGGAC 632  
Db 433 CTACACAGCCGACTAGTTATGATCAAGAGATTAATCTCTCAAGCAAGAACCTATGGGAC 492  
QY 693 CGAGCAGCTATGACAGCAGAGTATGATGATCAAGCAAGAGCTATGGGACAGCTCT 752  
Db 493 CGAGCAGCTATGACAGCAGAGTATGATGATCAAGCAAGAGCTATGGGACAGCTCT 552  
QY 753 CCACTAGTTACCAACCCCAACTGATCTTACAGCCAGCTCCAGTCAATATAGCCAC 812  
Db 553 CCACTAGTTACCAACCCCAACTGATCTTACAGCCAGCTCCAGTCAATATAGCCAC 612  
QY 813 AGAGCAGCAGCTATGAGGACAGAGTATCTTCCAGCAGACACCCCAAGTATGGGCTG 872  
Db 613 AGAGCAGCAGCTATGAGGACAGAGTATCTTCCAGCAGACACCCCAAGTATGGGCTG 672  
QY 873 TTTATGGGAGAGTCTGAGAGATTTTCCGGAACAGAGAGAACCCGAGCATGAGTGGCC 932  
Db 673 TTTATGGGAGAGTCTGAGAGATTTTCCGGAACAGAGAGAACCCGAGCATGAGTGGCC 732  
QY 933 CTGATTAACCGGGGACAGGGGAGAGGGGAGATTTGATCTGAGAGCATGAGAGTGGCC 992  
Db 733 CTGATTAACCGGGGACAGGGGAGAGGGGAGATTTGATCTGAGAGCATGAGAGTGGCC 792  
QY 993 GGGAGAGAGAGACCGCGTGGAGATGGGCGCTGAGAGCGAGGCTTCAATAGGCTG 1052  
Db 793 GGGAGAGAGAGACCGCGTGGAGATGGGCGCTGAGAGCGAGGCTTCAATAGGCTG 852  
QY 1053 GTGAGACCATGATGAGAGACAGATCTTGTATCTAGAGCCCACTGTGATCTCAAGATAG 1112  
Db 853 GTGAGACCATGATGAGAGACAGATCTTGTATCTAGAGCCCACTGTGATCTCAAGATAG 912  
QY 1113 ACTCTGACAAAGTGCATTTATGATACAGAGATTTAAATGACAGTGTACTTGAATGATC 1172  
Db 913 ACTCTGACAAAGTGCATTTATGATCAAGATTTAAATGACAGTGTACTTGAATGATC 972  
QY 1173 TGGCAGCTCTTTTAAAGCATGTGGGGTGTGTTAAGATGAACAGAGAACTGGGCAACCA 1232  
Db 973 TGGCAGCTCTTTTAAAGCATGTGGGGTGTGTTAAGATGAACAGAGAACTGGGCAACCA 1032  
QY 1233 TGATCCACATCTACTGAGCAAGAGAAACAGAAAGCCCAAGGCGATGCCAGTGTCT 1292  
Db 1033 TGATCCACATCTACTGAGCAAGAGAAACAGAAAGCCCAAGGCGATGCCAGTGTCT 1092  
QY 1293 ATGAGAACCAACCACTGACAGAGCTGCGGTGAGATGTTGATGGGAAAGATTTTCAAG 1352  
Db 1093 ATGAGAACCAACCACTGACAGAGCTGCGGTGAGATGTTGATGGGAAAGATTTTCAAG 1152  
QY 1353 GAGCAAACTTTAAAGTCTCTCTGCTGAGAAAGCTCCCAATGAATGAGGGGTG 1412  
Db 1153 GAGCAAACTTTAAAGTCTCTCTGCTGAGAAAGCTCCCAATGAATGAGGGGTG 1212  
QY 1413 GTCTGCAACCCGCTGAGAGGAGAGAGTGCACCAACCACTCCGAGAGCTCCAGAGAGCC 1472  
Db 1213 GTCTGCAACCCGCTGAGAGGAGAGAGTGCACCAACCACTCCGAGAGCTCCAGAGAGCC 1272  
QY 1473 CAGAGGTCTCTGGGAGACCCATGGGTGCAATGGAGGCGGTGAGAGATGAGAGAGCT 1532  
Db 1273 CAGAGGTCTCTGGGAGACCCATGGGTGCAATGGAGGCGGTGAGAGATGAGAGAGCT 1332  
QY 1533 TCCTCTCAAGAGACCCCGGGGTTCCTGAGAGAGAACCTCTGAGAGAGAACTCCAGC 1592  
Db 1333 TCCTCTCAAGAGACCCCGGGGTTCCTGAGAGAGAACCTCTGAGAGAGAACTCCAGC 1392  
QY 1593 ACCGAGCTGAGAGCTGAGAGTGTCCCAATCCGGGTGTGAGAAACCAAGATCTTGGCTGAG 1652

Db 1393 ACCGAGCTGAGAGCTGAGAGTGTCCCAATCCGGGTGTGAGAAACCAAGATCTTGGCTGAG 1452  
QY 1653 GAGCAGAGTGCACCACTGTGAGAGGCCCCCAAGGCTTCTCTCCGCAACCTTTC 1712  
Db 1453 GAGCAGAGTGCACCACTGTGAGAGGCCCCCAAGGCTTCTCTCCGCAACCTTTC 1512  
QY 1713 CGCCCCGGGTGTGATCTGAGCAGAGTGGCCCTGTGAGCATGCGGGAGAGAGAGGTG 1772  
Db 1513 CGCCCCGGGTGTGATCTGAGCAGAGTGGCCCTGTGAGCATGCGGGAGAGAGAGGTG 1572  
QY 1773 GCCTCATGAGATCGTGTGATCCCGTGAATGTTCAAGGTGGCCGTGTGAGACAGAG 1832  
Db 1573 GCCTCATGAGATCGTGTGATCCCGTGAATGTTCAAGGTGGCCGTGTGAGACAGAG 1632  
QY 1833 GTGGCTTCGCTGTGGCCCGGGGATGAGACCGAGGTGCTTGTGTGAGAGAGAGGTG 1892  
Db 1633 GTGGCTTCGCTGTGGCCCGGGGATGAGACCGAGGTGCTTGTGTGAGAGAGAGGTG 1692  
QY 1893 GCGCTGGGGGGCCCCCTGAGACCTTGTGAGAGAGAGAGAGAGAGAGAGAGAGGTG 1952  
Db 1693 GCGCTGGGGGGCCCCCTGAGACCTTGTGAGAGAGAGAGAGAGAGAGAGAGAGGTG 1752  
QY 1953 GAGGACCTGAGAAATGATGATTAAGGCGAGACCCGTGAGAGCCGAGATCGGCTTACT 2012  
Db 1753 GAGGACCTGAGAAATGATGATTAAGGCGAGACCCGTGAGAGCCGAGATCGGCTTACT 1812  
QY 2013 AGATGAGAGACCCCGGAGAGCTGATGATGATGATGATGATGATGATGATGATGAT 2072  
Db 1813 AGATGAGAGACCCCGGAGAGCTGATGATGATGATGATGATGATGATGATGATGAT 1872  
QY 2073 GTTTAAATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTAAT 2132  
Db 1873 GTTTAAATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTAAT 1932  
QY 2133 CTGTACTTATGATTTTTCACATTTGTGAGAGAACTTTAAACAACTTAAATGCT 2188  
Db 1933 CTGTACTTATGATTTTTCACATTTGTGAGAGAACTTTAAACAACTTAAATGCT 1988

RESULT 8  
US-10-425-115-182496  
; Sequence 182496, Application US/10425115  
; Publication No. US20040214272A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa, Thomas J.  
; APPLICANT: Kovacic, David K.  
; APPLICANT: Zhou, Yihua  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
; FILE REFERENCE: 38-21(53222)B  
; CURRENT APPLICATION NUMBER: US/10/425,115  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 369326  
; SEQ ID NO 182496  
; LENGTH: 763  
; TYPE: DNA  
; ORGANISM: Zea mays  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: (1)-(763)  
; OTHER INFORMATION: unsure at all n locations  
; FEATURE:  
; OTHER INFORMATION: Clone ID: MR14577\_98018C.1  
US-10-425-115-182496

Query Match 26.4%; Score 631; DB 18; Length 763;  
Best Local Similarity 92.8%; Pred. No. 9.9e-151;  
Matches 661; Conservative 0; Mismatches 51; Indels 0; Gaps 0;  
QY 36 GAGAGAAATGCGTCCAGGATTAAGTACTATAGCCAGAGTGCAGCCAGAGGAGCT 95



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 20, 2005, 04:03:23 ; Search time 7059.46 Seconds  
(without alignments)  
12886.762 Million cell updates/sec

Title: US-10-791-017A-1

Perfect score: 2390  
Sequence: 1 agagagagagagagagcgttgag.....taaaaaaaaaaaaaaaaaa 2390

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*  
1: gb\_esc1:\*  
2: gb\_esc2:\*  
3: gb\_hcc:\*  
4: gb\_esc3:\*  
5: gb\_esc4:\*  
6: gb\_esc5:\*  
7: gb\_esc6:\*  
8: gb\_g881:\*  
9: gb\_g882:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2314	96.8	2314	3	CR608375 full-length
2	2157	90.3	2157	3	CR604952 full-length
3	2146	89.8	2236	3	CR604639 full-length
4	2144	89.7	2159	3	CR625247 full-length
5	2113	88.4	2126	3	CR610888 full-length
6	2103	88.0	2103	3	CR619493 full-length
7	2083	87.2	2083	3	CR608666 full-length
8	2080	87.0	2093	3	CR594868 full-length
9	2073	86.7	2073	3	CR593724 full-length
10	2057	86.1	2070	3	CR602561 full-length
11	1979.8	82.8	2373	3	AK034755 Mus muscu
12	1860	77.8	2107	3	AK019460 Mus muscu
13	1769.8	74.1	2269	3	AK049743 Mus muscu
14	1094.4	45.8	1379	3	AK014366 Mus muscu
15	989.8	41.4	1049	5	BX325713 BX325713
16	976.8	40.9	1008	5	BX353313 BX353313
17	959.4	40.1	1020	5	BX387336 BX387336
18	956.8	40.0	1021	5	BX464988 BX464988
19	944.4	39.5	1062	4	BM473356 AGENCOURT
20	935.8	39.2	984	5	BX443496 BX443496
21	935.4	39.1	1013	5	BX367687 BX367687
22	930.2	38.9	993	5	BX462635 BX462635
23	928.6	38.9	998	5	BX379802 BX379802
24	928.4	38.8	1045	5	BM909438 AGENCOURT

25	928.2	38.8	1020	5	BM927718 AGENCOURT
26	926.4	38.8	1024	5	BQ062138 AGENCOURT
27	925.2	38.7	1079	5	BX449774 BX449774
28	923.2	38.6	1057	5	BX353312 BX353312
29	920.8	38.5	1031	4	BM477969 AGENCOURT
30	920.2	38.5	1056	5	BQ059127 AGENCOURT
31	918.2	38.4	1058	4	BM802924 AGENCOURT
32	917	38.4	1024	4	BM468711 AGENCOURT
33	916.6	38.4	1009	5	BX421725 BX421725
34	912.2	38.1	1068	5	BX382091 BX382091
35	910.6	38.1	971	5	BX407826 BX407826
36	908.4	38.0	962	5	BX443095 BX443095
37	907	37.9	1029	5	BQ059077 AGENCOURT
38	907	37.9	1048	5	BX381313 BX381313
39	906.6	37.9	962	5	BX407870 BX407870
40	904.6	37.8	931	5	BX429003 BX429003
41	904.6	37.8	1068	5	BX449773 BX449773
42	904	37.8	948	5	BX424687 BX424687
43	904	37.8	1041	5	BX420692 BX420692
44	903.8	37.8	1040	5	BX399070 BX399070
45	903.4	37.8	1027	5	BX428312 BX428312

#### ALIGNMENTS

RESULT 1	CR608375	2314 bp	mrna	linear	HTC 21-JUL-2004
LOCUS	full-length cDNA clone CS0D1039J24 of Placenta Cot 25-normalized				
DEFINITION	of Homo sapiens (human) .				
ACCESSION	CR608375.1	GI:50489182			
VERSION	HTC; CNSTL cDNA.				
KEYWORDS	Homo sapiens (human)				
SOURCE	Homo sapiens				
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
REFERENCE	1 (bases 1 to 2314)				
AUTHORS	Li, W.B., Gruber, C., Jesse, J., and Polayes, D.				
TITLE	Full-length cDNA libraries and normalization				
JOURNAL	Unpublished				
REMARK	Contact : Feng Liang Email : fliang@life.techn.com URL : http://fulllength.invitrogen.com/ Invitrogen Corporation 1600 Faraday Avenue 2 (bases 1 to 2314)				
REFERENCE	Genoscope.				
AUTHORS	Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage :				
TITLE	BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr				
JOURNAL	- Web : www.genoscope.cns.fr)				
COMMENT	1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies, a division of Invitrogen.				
FEATURES	Location/Qualifiers				
source	1. 2314				
	/organism="Homo sapiens"				
	/mol_type="mrna"				
	/db_xref="taxon:9606"				
	/clone="CS0D1039J24"				
	/issue_type="Placenta Cot 25-normalized"				
	/plasmid="pCMVSPORT_6"				
ORIGIN					
Query Match	96.8%; Score 2314; DB 3; Length 2314;				
Best Local Similarity	100.0%; Pred. No. 0;				
Matches 2314; Conservative	0; Mismatches 0; Indels 0; Gaps 0;				
QY	4 GGGAGACGAGCGTTGAGAGACGAGAGAGAGAGAAATGGCTCCACGATTACAG 63				
DB	1 GGGAGACGAGCGTTGAGAGACGAGAGAGAGAGAGAAATGGCTCCACGATTACAG 60				

QY 64 TACCTTAAGCAAGCTGACGAGCGGCTACAGTGTCTTAACCGCCCAAGCCCTCA 123  
DB 61 TACCTTAAGCAAGCTGACGAGCGGCTACAGTGTCTTAACCGCCCAAGCCCTCA 120  
QY 124 AGGATATGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCA 183  
DB 121 AGGATATGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCA 180  
QY 184 CACTGTATGCAAGCTATACCAAGGCTGAGCAAGCAAGCTATGAGCAAGCGCTATG 243  
DB 181 CACTGTATGCAAGCTATACCAAGGCTGAGCAAGCAAGCTATGAGCAAGCGCTATG 240  
QY 244 AACTTCTTATGCAAGCTATACCAAGGCTGAGCAAGCAAGCTATGAGCAAGCGCTATG 303  
DB 241 AACTTCTTATGCAAGCTATACCAAGGCTGAGCAAGCAAGCTATGAGCAAGCGCTATG 300  
QY 304 CAGGCTGTCAGAGGATGAGCAAGCTGATGATGATGATGATGATGATGATGATGATG 363  
DB 301 CAGGCTGTCAGAGGATGAGCAAGCTGATGATGATGATGATGATGATGATGATGATG 360  
QY 364 CAGGCTGTCAGAGGATGAGCAAGCTGATGATGATGATGATGATGATGATGATGATG 423  
DB 361 CAGGCTGTCAGAGGATGAGCAAGCTGATGATGATGATGATGATGATGATGATGATG 420  
QY 424 TGGGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCA 483  
DB 421 TGGGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCA 480  
QY 484 GACTAGTCAACTCAATCTAGCAAGAGGAGTCAAGCAAGCAAGCAAGCAAGCAAGCA 543  
DB 481 GACTAGTCAACTCAATCTAGCAAGAGGAGTCAAGCAAGCAAGCAAGCAAGCAAGCA 540  
QY 544 GAGTAACTACAGTTATCCCAAGGTAAGCTGAGAGTACCCATGAGCAAGCAAGCA 603  
DB 541 GAGTAACTACAGTTATCCCAAGGTAAGCTGAGAGTACCCATGAGCAAGCAAGCAAG 600  
QY 604 TCCATCTTACCTCTTACCAAGCTATCTCTTACCAAGCAAGCAAGCAAGCAAGCAAG 663  
DB 601 TCCATCTTACCTCTTACCAAGCTATCTCTTACCAAGCAAGCAAGCAAGCAAGCAAG 660  
QY 664 TTAATCTCAGCAAGCAAGCTATGAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 723  
DB 661 TTAATCTCAGCAAGCAAGCTATGAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 720  
QY 724 TCAACAAAGCAAGCTATGAGCAAGCTTCCCAAGTATGATGATGATGATGATGATG 783  
DB 721 TCAACAAAGCAAGCTATGAGCAAGCTTCCCAAGTATGATGATGATGATGATGATG 780  
QY 784 CAGGCAAGCTCAGAGTCAATATGAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 843  
DB 781 CAGGCAAGCTCAGAGTCAATATGAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 840  
QY 844 CCGACAGAGCAAGCCCAAGTATGAGTGTGTTATGAGCAAGCAAGTCTGAGAGATTTTC 903  
DB 841 CCGACAGAGCAAGCCCAAGTATGAGTGTGTTATGAGCAAGCAAGTCTGAGAGATTTTC 900  
QY 904 ACCAGAGAGAACCGGAGCAAGTATGAGTGTGTTATGAGCAAGCAAGTCTGAGAGATTT 963  
DB 901 ACCAGAGAGAACCGGAGCAAGTATGAGTGTGTTATGAGCAAGCAAGTCTGAGAGATTT 960  
QY 964 TGAATCGTGAAGGCAAGTATGAGTGTGTTATGAGCAAGCAAGTCTGAGAGATTTTC 1023  
DB 961 TGAATCGTGAAGGCAAGTATGAGTGTGTTATGAGCAAGCAAGTCTGAGAGATTTTC 1020  
QY 1024 TGGAGAGAGAGTGTGTTATGAGTGTGTTATGAGCAAGCAAGTCTGAGAGATTTTC 1083  
DB 1021 TGGAGAGAGAGTGTGTTATGAGTGTGTTATGAGCAAGCAAGTCTGAGAGATTTTC 1080  
QY 1084 TCTAGAGCCCACTGTATGATGAGTGAAGCAAGTCTGAGCAAGTCTGAGCAAGTCTGAG 1143  
DB 1081 TCTAGAGCCCACTGTATGATGAGTGAAGCAAGTCTGAGCAAGTCTGAGCAAGTCTGAG 1140

QY 1144 ATTAATATGCAAGTATGATCTATGATGATGATGATGATGATGATGATGATGATGAT 1203  
DB 1141 ATTAATATGCAAGTATGATCTATGATGATGATGATGATGATGATGATGATGATGAT 1200  
QY 1204 TAAAGTGAACAAAGAACTGGGCAAGCCATGATGATGATGATGATGATGATGATGATG 1263  
DB 1201 TAAAGTGAACAAAGAACTGGGCAAGCCATGATGATGATGATGATGATGATGATGATG 1260  
QY 1264 AAAAGCCAAAGGCGATGCGCAAGTGTCTATGATGATGATGATGATGATGATGATGAT 1323  
DB 1261 AAAAGCCAAAGGCGATGCGCAAGTGTCTATGATGATGATGATGATGATGATGATGAT 1320  
QY 1324 GGAATGTTGATGAGGAAAGATTTTCAAGGAGCAAACTTAAAGTCTGCTGCTGAGAA 1383  
DB 1321 GGAATGTTGATGAGGAAAGATTTTCAAGGAGCAAACTTAAAGTCTGCTGCTGAGAA 1380  
QY 1384 GAAAGCTTCAATGAACAGTATGAGGAGTGTGTCACACCCGTGAGGAGCAAGCAAGT 1443  
DB 1381 GAAAGCTTCAATGAACAGTATGAGGAGTGTGTCACACCCGTGAGGAGCAAGCAAGT 1440  
QY 1444 ACCACCACTGCGGAGAGTCCAGAGAGCCCAAGAGTCTGAGGAGCAAGTGTGATGAT 1503  
DB 1441 ACCACCACTGCGGAGAGTCCAGAGAGCCCAAGAGTCTGAGGAGCAAGTGTGATGAT 1500  
QY 1504 GGGAGGCGGTGAGAGATGAGAGAGCTTCCCTCCAAAGAGCAAGCCGAGGAGTCCAGAG 1563  
DB 1501 GGGAGGCGGTGAGAGATGAGAGAGCTTCCCTCCAAAGAGCAAGCCGAGGAGTCCAGAG 1560  
QY 1564 GAAAGCTTCTGAGAGAGAGAGAGTCCAGAGCAAGAGTGAAGCTGAGAGTGTCCAAAT 1623  
DB 1561 GAAAGCTTCTGAGAGAGAGAGAGTCCAGAGCAAGAGTGAAGCTGAGAGTGTCCAAAT 1620  
QY 1624 GGGTGTGAGAAACCAAGAACTTGGCTGAGAGAAAGAGTCAAGCAAGTGAAGGCCCCAA 1683  
DB 1621 GGGTGTGAGAAACCAAGAACTTGGCTGAGAGAAAGAGTCAAGCAAGTGAAGGCCCCAA 1680  
QY 1684 GCTGAGAGGCTTCTGCTCCGCAAGCTTCCGCTGAGGAGTGTGATGATGAGCAAGAGT 1743  
DB 1681 GCTGAGAGGCTTCTGCTCCGCAAGCTTCCGCTGAGGAGTGTGATGATGAGCAAGAGT 1740  
QY 1744 CCTGAGTGAATGAGGAGAGAGAGAGTGTGATGATGATGATGATGATGATGATGATG 1803  
DB 1741 CCTGAGTGAATGAGGAGAGAGAGTGTGATGATGATGATGATGATGATGATGATGATG 1800  
QY 1804 GTTCAAGAGTGTGCTGAGAGAGAGAGTGTGATGATGATGATGATGATGATGATGATG 1863  
DB 1801 GTTCAAGAGTGTGCTGAGAGAGAGAGTGTGATGATGATGATGATGATGATGATGATG 1860  
QY 1864 AGGTGCTTGTGAGAGAGAGAGAGTGTGATGATGATGATGATGATGATGATGATGATG 1923  
DB 1861 AGGTGCTTGTGAGAGAGAGAGAGTGTGATGATGATGATGATGATGATGATGATGATG 1920  
QY 1924 ACAGATGAGAGAGAGAGAGAGAGAGAGTGTGATGATGATGATGATGATGATGATGATG 1983  
DB 1921 ACAGATGAGAGAGAGAGAGAGAGAGAGTGTGATGATGATGATGATGATGATGATGATG 1980  
QY 1984 CCGTCAAGAGAGAGAGAGAGAGAGAGTGTGATGATGATGATGATGATGATGATGATG 2043  
DB 1981 CCGTCAAGAGAGAGAGAGAGAGAGAGTGTGATGATGATGATGATGATGATGATGATG 2040  
QY 2044 TACCAATTTATTTTAAACAGAAATGTTTAAATTTTAAATTTTAAATTTTAAATTT 2103  
DB 2041 TACCAATTTATTTTAAACAGAAATGTTTAAATTTTAAATTTTAAATTTTAAATTT 2100  
QY 2104 TTGGCAACAATTTATGATTTATCTTGTCTGATGATGATGATGATGATGATGATGATG 2163  
DB 2101 TTGGCAACAATTTATGATTTATCTTGTCTGATGATGATGATGATGATGATGATGATG 2160  
QY 2164 GAAACATTTAAACAAAGTAAATGATGATGATGATGATGATGATGATGATGATGATG 2223  
DB 2161 GAAACATTTAAACAAAGTAAATGATGATGATGATGATGATGATGATGATGATGATG 2220  
QY 2224 AATGATGTTTAAAGCTTTTAAACAAATGAGAAACCTTGTGAGAGATGATGATGATG 2283

Db 2221 AATGTTGTTTAAAGACTTTTAAAGAGGAAACCCCTTGTGAGCATGCTCAGATTCATTGTG 2280  
QY 2284 GAGAACCAAGAGGGCCCTTAACTGTAAACATGT 2217  
Db 2281 GAGAACCAAGAGGGCCCTTAACTGTAAACATGT 2214

RESULT 2  
CR604952 2157 bp mRNA linear HTC 21-JUL-2004  
LOCUS full-length cDNA clone CS0D1010Y112 of Placenta Cot 25-normalized  
DEFINITION of Homo sapiens (human).  
ACCESSION CR604952.1 GI:50485759  
VERSION CR604952.1  
KEYWORDS HTC; CNSLT; cDNA.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 2157)  
AUTHORS Li, W.B., Gruber, C., Jesssee, J. and Polayes, D.  
TITLE Full-length cDNA libraries and normalization  
JOURNAL Unpublished  
REMARK Contact : Feng Liang Email : fliang@lifetech.com URL :  
http://fulllength.invitrogen.com/Invitrogen Corporation 1600  
Faraday Avenue  
2 (bases 1 to 2157)  
REFERENCE Genoscope.  
AUTHORS Direct Submission  
JOURNAL Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage :  
BP 131 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)  
COMMENT - Web : www.genoscope.cns.fr)  
1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime  
end enriched, double-strand cDNA was digested with Not I and cloned  
into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library  
was normalized. Library was constructed by Life Technologies, a  
division of Invitrogen.  
FEATURES  
source location/qualifiers  
1..2157  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="CS0D1010Y112"  
/issue\_type="Placenta Cot 25-normalized"  
/plasmid="pCMVSPORT\_6"

ORIGIN  
Query Match 90.3%; Score 2157; DB 3; Length 2157;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 2157; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 GGGAGACGAGCGTTGAGAGAACGAGAGAGAGAGAAATGGCGTCCACGGATTTCAG 63  
Db 1 GGGAGACGAGCGTTGAGAGAACGAGAGAGAGAGAAATGGCGTCCACGGATTTCAG 60  
QY 64 TACCTATAGCCAAAGCTGACGCGAGCGAGGCTACAGTGTCTTACACCGCCAGCCACTCA 123  
Db 61 TACCTATAGCCAAAGCTGACGCGAGCGAGGCTACAGTGTCTTACACCGCCAGCCACTCA 120  
QY 124 AGAGTATGCAAGACCCAGGAGATATGGGCAACAAGCTATGGAACCTATGACAGCC 183  
Db 121 AGAGTATGCAAGACCCAGGAGATATGGGCAACAAGCTATGGAACCTATGACAGCC 180  
QY 184 CACGTATGTCAGCTATACCAAGGCTCAGACCACTGCAACCTATGGGAGACCGGCTATGC 243  
Db 181 CACTGATGTCAGCTATACCAAGGCTCAGACCACTGCAACCTATGGGAGACCGGCTATGC 240  
QY 244 AACTTTCTTATGACAGGCTCCACTGTTATTAATCTCAACTGCCCCCGAGGATACAG 303  
Db 241 AACTTTCTTATGACAGGCTCCACTGTTATTAATCTCAACTGCCCCCGAGGATACAG 300  
QY 304 CAGGCTGTCCAGGGGATATGGCACTGGTGTCTTATGATACCAACCACTGCTACAGTACAC 363

Db 301 CAGGCTGTCCAGGGGATATGGCACTGGTGTCTTATGATACCAACCACTGCTACAGTACAC 360  
QY 364 CACCCAGGCTCTCTATGACAGCTCAGTCTGATATGGCACTCAGCTGCTTATCCAGCTTA 423  
Db 361 CACCCAGGCTCTCTATGACAGCTCAGTCTGATATGGCACTCAGCTGCTTATCCAGCTTA 420  
QY 424 TGGGACAGGCTCCAGGACCTGCACTGCAACCGCAGATGGAACCAAGCCCACTGA 483  
Db 421 TGGGACAGGCTCCAGGACCTGCACTGCAACCGCAGATGGAACCAAGCCCACTGA 480  
QY 484 GACTATGCAACCTCAATCTTATGACAGGGGGTTTCAACAGCCGAGCTAGATATGAGCA 543  
Db 481 GACTATGCAACCTCAATCTTATGACAGGGGGTTTCAACAGCCGAGCTAGATATGAGCA 540  
QY 544 GAGTAACTACAGTTATCCCAAGGATCTGGAGCTAACCCAGTCAAGCCAGTCACTGACC 603  
Db 541 GAGTAACTACAGTTATCCCAAGGATCTGGAGCTAACCCAGTCAAGCCAGTCACTGACC 600  
QY 604 TCCATCTTACCTCTTACCAAGCTATTCCTTATACAGCCGAGCTATGATATGACAGAG 663  
Db 601 TCCATCTTACCTCTTACCAAGCTATTCCTTATACAGCCGAGCTATGATATGACAGAG 660  
QY 664 TTACTCTCAGAGAACACCTATGAGGCAACCGAGAGCTATGACAGAGAGTACTATG 723  
Db 661 TTACTCTCAGAGAACACCTATGAGGCAACCGAGAGCTATGAGAGAGAGTACTATG 720  
QY 724 TCAACAAAGAGCTATGAGGAGAGGCTCCCACTAGTTAACCCCAACCTGATCTTA 783  
Db 721 TCAACAAAGAGCTATGAGGAGAGGCTCCCACTAGTTAACCCCAACCTGATCTTA 780  
QY 784 CAGCCAGCTCTCAATGATATATGCGCAACAGAGAGAGCTACGGGAGAGAGTTCATT 843  
Db 781 CAGCCAGCTCTCAATGATATATGCGCAACAGAGAGAGCTACGGGAGAGAGTTCATT 840  
QY 844 CCGACAGAGACACCCAGTATGAGGAGTGTATGAGGAGAGAGTCTGAGAGATTTTCGG 903  
Db 841 CCGACAGAGACACCCAGTATGAGGAGTGTATGAGGAGAGAGTCTGAGAGATTTTCGG 900  
QY 904 ACCAGAGAGAACCGAGAGCTAGTGTGCTCTGATTAACCGGGAGAGGAGAGGAGATT 963  
Db 901 ACCAGAGAGAACCGAGAGCTAGTGTGCTCTGATTAACCGGGAGAGGAGAGGAGATT 960  
QY 964 TGATCTGAGGAGCTATGACAGAGTGGCGGGAGAGAGAGCGGTGGAATGGGACGGC 1023  
Db 961 TGATCTGAGGAGCTATGACAGAGTGGCGGGAGAGAGAGCGGTGGAATGGGACGGC 1020  
QY 1024 TGGAGAGCGAGGCTTCAATTAAGCTGTGTGAACCATGGAAGGAAGACAGATCTGA 1083  
Db 1021 TGGAGAGCGAGGCTTCAATTAAGCTGTGTGAACCATGGAAGGAAGACAGATCTGA 1080  
QY 1084 TCTAGGCCCCACTGTAGATCTCAGATGAAGACTCTGCAACAAGTGAATTTATGACAGG 1143  
Db 1081 TCTAGGCCCCACTGTAGATCTCAGATGAAGACTCTGCAACAAGTGAATTTATGACAGG 1140  
QY 1144 ATTAAATGACAGTGTGATCTTATGATATCTGCGAAGCTTTTAAAGAGAGTGGGGTTGT 1203  
Db 1141 ATTAAATGACAGTGTGATCTTATGATATCTGCGAAGCTTTTAAAGAGAGTGGGGTTGT 1200  
QY 1204 TAAAGATGAACAAGAACTGGGCAACCGATGATCCCATCTTACTGGAACAAGAAACAG 1263  
Db 1201 TAAAGATGAACAAGAACTGGGCAACCGATGATCCCATCTTACTGGAACAAGAAACAG 1260  
QY 1264 AAAGCCCAAGGAGATGCGACAGATGCTTATGAGAACCCACCACTGCGCAAGGCTGCGT 1323  
Db 1261 AAAGCCCAAGGAGATGCGACAGATGCTTATGAGAACCCACCACTGCGCAAGGCTGCGT 1320  
QY 1324 GGAATGGTTGATGAGAAAGATTTTCAAGGAGCAAACTTAAAGTCTCCTTCTCGGAA 1383  
Db 1321 GGAATGGTTGATGAGAAAGATTTTCAAGGAGCAAACTTAAAGTCTCCTTCTCGGAA 1380  
QY 1384 GAAGCTTCCAAATGAAGATATGGGGGTGTCTGCAACCCCGTGAAGGAGAGGATGCC 1443



Db 721 CAGCTCCCACTAGTACCCACCCCAACTGATCTTACAGCCAGCTCCAGTCAATAT 780  
 QY 806 AGCCACAGACGACGACCTACGGGACAGAGATTCAATCCGACAGAACCAACCCAGTAGC 865  
 Db 781 AGCCACAGACGACGACCTACGGGACAGAGATTCAATCCGACAGAACCAACCCAGTAGC 840  
 QY 866 ATGGGTGTTATATGGGACGAGAGTCTGAGAGATTTTCCGACACAGAGAGAACCCGAGCATG 925  
 Db 841 ATGGGTGTTATATGGGACGAGAGTCTGAGAGATTTTCCGACACAGAGAGAACCCGAGCATG 900  
 QY 926 AGTGGCCCTGATAACCGGGGACAGGGGAGAGAGGGGATTTGATCGTGGAGCAGTAGCAGA 985  
 Db 901 AGTGGCCCTGATAACCGGGGACAGGGGAGAGAGGGGATTTGATCGTGGAGCAGTAGCAGA 960  
 QY 986 GGTGGGGGGGAGAGAGACCGGTGGAATGGGACGCCCTGAGAGACGAGGTGCTCAAT 1045  
 Db 961 GGTGGGGGGGAGAGAGACCGGTGGAATGGG---GCCCTGAGAGACGAGGTGCTCAAT 1017  
 QY 1046 AAGCTGTGGAGACCCATGATGAGAGACGAGTCTTATCTAGGCCCACTGTATGATCCA 1105  
 Db 1018 AAGCTGTGGAGACCCATGATGAGAGACGAGTCTTATCTAGGCCCACTGTATGATCCA 1077  
 QY 1106 GATGAGACTCTGACCAACAGTGCATTTATGACAGAGATTAATGACAGTGTACTCTA 1165  
 Db 1078 GATGAGACTCTGACCAACAGTGCATTTATGACAGAGATTAATGACAGTGTACTCTA 1137  
 QY 1166 GATGATCTGCAACCTTTTAAAGACAGTGTGGGTTTAAATGAACAAGAACTGGG 1225  
 Db 1138 GATGATCTGCAACCTTTTAAAGACAGTGTGGGTTTAAATGAACAAGAACTGGG 1197  
 QY 1226 CAACCCATGATCACTACCTACCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1285  
 Db 1198 CAACCCATGATCACTACCTACCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1257  
 QY 1286 GTGTCTATGAGAGACCCCACTGCAAGAGCTGCCGTGAAATGTTGATGGAGAAAT 1345  
 Db 1258 GTGTCTATGAGAGACCCCACTGCAAGAGCTGCCGTGAAATGTTGATGGAGAAAT 1317  
 QY 1346 TTTCAAGAGAGCAAACTTAAAGTCTCCCTTGTCTGAGAGAGAGAGAGAGAGAGAGAG 1405  
 Db 1318 TTTCAAGAGAGCAAACTTAAAGTCTCCCTTGTCTGAGAGAGAGAGAGAGAGAGAGAG 1377  
 QY 1406 CGGGGTGTGTGCAACCCCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1465  
 Db 1378 CGGGGTGTGTGCAACCCCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1437  
 QY 1466 GAGAGCCCAAGAGAGTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1525  
 Db 1438 GAGAGCCCAAGAGAGTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1497  
 QY 1526 GAGAGCTTCCCTCCAG 1585  
 Db 1498 GAGAGCTTCCCTCCAG 1557  
 QY 1586 GTCCACAGACCGAGCTGAG 1645  
 Db 1558 GTCCACAGACCGAGCTGAG 1617  
 QY 1646 GCTGTGAGAGACAGAGTCAACAGTGTAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1705  
 Db 1618 GCTGTGAGAGACAGAGTCAACAGTGTAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1677  
 QY 1706 CCGTTTCCGCCCCCGGGGTGTGATCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1765  
 Db 1678 CCGTTTCCGCCCCCGGGGTGTGATCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1737  
 QY 1766 AGAGGTGTGCTCATGATCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1825  
 Db 1738 AGAGGTGTGCTCATGATCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1797  
 QY 1826 GACAGAGTGTGCTTCCGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1885  
 Db 1798 GACAGAGTGTGCTTCCGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1857

QY 1886 CGAGTGTGCTTGTGAG 1945  
 Db 1858 CGAGTGTGCTTGTGAG 1879  
 QY 1946 GAGAGTGTGAG 2005  
 Db 1880 -----AGGAG 1910  
 QY 2006 CCTACTAGATGAG 2065  
 Db 1911 CCTACTAGATGAG 1970  
 QY 2066 AGAAATGTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2125  
 Db 1971 AGAAATGTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2030  
 QY 2126 TCCTGTGTGCTACTAGTATGATTTTCAACATTTGTGAGAGAGAGAGAGAGAGAGAGAG 2185  
 Db 2031 TCCTGTGTGCTACTAGTATGATTTTCAACATTTGTGAGAGAGAGAGAGAGAGAGAGAG 2090  
 QY 2186 GGTAGTGTGCGAGATTTTCTTCTCTTTTAAATGAGTGTGTTAAGACTTAAAC 2245  
 Db 2091 GGTAGTGTGCGAGATTTTCTTCTCTTTTAAATGAGTGTGTTAAGACTTAAAC 2150  
 QY 2246 AATGGAG 2305  
 Db 2151 AATGGAG 2210  
 QY 2306 CTGTAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2331  
 Db 2211 CTGTAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2236

RESULT 4  
 CR625247  
 LOCUS  
 DEFINITION  
 full-length cDNA clone CSDB005IF11 of Placenta of Homo sapiens (human).  
 ACCESSION  
 CR625247.1 GI:50506054  
 VERSION  
 HTC; CDSLT CDNA.  
 KEYWORDS  
 Homo sapiens (human)  
 SOURCE  
 Homo sapiens  
 ORGANISM  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE  
 AUTHORS  
 Li W.B., Gruber, C., Jesse, J. and Polyes, D.  
 TITLE  
 Full-length cDNA libraries and normalization  
 JOURNAL  
 Unpublished  
 REMARK  
 Contact : Feng Liang Email : fliang@life.techn.com URL : http://fulllength.invitrogen.com/ Invitrogen Corporation 1600 Faraday Avenue  
 2 (bases 1 to 2159)  
 REFERENCE  
 AUTHORS  
 Genoscope.  
 TITLE  
 Direct Submission  
 JOURNAL  
 Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE [E-mail : seqref@genoscope.cns.fr - web : www.genoscope.cns.fr]  
 COMMENT  
 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies, a division of Invitrogen.  
 FEATURES  
 location/Qualifiers  
 1..2159  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="CSDB005IF11"  
 /issue\_type="Placenta"  
 /plasmid="pCMVSPORT\_6"  
 ORIGIN







Db 2100 ATGTGGCCACACATTATATTCCTTCTGTCTTAGATTTTCCACATTGT 2159  
RESULT 5  
CR610888  
LOCUS  
DEFINITION  
CR610888 2126 bp mRNA linear HTC 21-JUL-2004  
full-length cDNA clone CS0DF037Y107 of Fetal brain of Homo sapiens  
(human).  
ACCESSION  
CR610888 GI:50491695  
VERSION  
CR610888.1  
KEYWORDS  
HTC; CNSLT cDNA.  
SOURCE  
Homo sapiens (human)  
ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE  
1 (bases 1 to 2126)  
Li, W. B., Gruber, C., Jesses, J. and Polayes, D.  
Full-length cDNA libraries and normalization  
Unpublished  
JOURNAL  
Contact : Peng Liang Email : filiang@lifetech.com URL :  
http://fulllength.invitrogen.com/Invitrogen Corporation 1600  
Faraday Avenue  
Genoscope.  
2 (bases 1 to 2126)  
Genoscope.  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
REMARK  
COMMENT  
Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage :  
BP 131 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr  
- web : www.genoscope.cns.fr)  
1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime  
end enriched, double-strand cDNA was digested with Not I and cloned  
into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library  
was normalized. Library was constructed by Life Technologies, a  
division of Invitrogen.  
FEATURES  
source  
1. .2126  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="CS0DF037Y107"  
/issue\_type="Fetal brain"  
/plasmid="pCMVSPORT\_6"  
ORIGIN  
Query Match 88.4%; Score 2113; DB 3; Length 2126;  
Best Local Similarity 99.9%; Pred. No. 0;  
Matches 2126; Conservative 0; Mismatches 0; Indels 3; Gaps 1;  
QY 56 GATTACAGTACCTATAGCCAGGCTGACGCGCAGAGGCTTACAGCGCCAG 115  
DB 1 GATTACAGTACCTATAGCCAGGCTGACGCGCAGAGGCTTACAGCGCCAG 60  
QY 116 CCCACTCAAGATATGACAGACCAACCAGCATATGGGCAAAAGCTATGAACTTAT 175  
DB 61 CCACCTCAAGATATGACAGACCAACCAGCATATGGGCAAAAGCTATGAACTTAT 120  
QY 176 GGAAGAGCCCACTGATGTACCTATACCCAGGCTTCAAGCACTGCAACTATGGGCAAGC 235  
DB 121 GGAAGAGCCCACTGATGTACCTATACCCAGGCTTCAAGCACTGCAACTATGGGCAAGC 180  
QY 236 GCGTATGCAACTTCTATGAGCAGCGCTCCCACTGTTATCTACTCAATGCGCCCCAG 295  
DB 181 GCGTATGCAACTTCTATGAGCAGCGCTCCCACTGTTATCTACTCAATGCGCCCCAG 240  
QY 296 GCATACAGCAGCGCTGTCAGGGGGTATGCACTGTGCTTATGATACCACTGTCTAC 355  
DB 241 GCATACAGCAGCGCTGTCAGGGGGTATGCACTGTGCTTATGATACCACTGTCTAC 300  
QY 356 GTTACCAACCAACCAAGGCTCTTATGAGCTCACTGTGATATGCACTGAGCTGCTTAT 415  
DB 301 GTTACCAACCAACCAAGGCTCTTATGAGCTCACTGTGATATGCACTGAGCTGCTTAT 360  
QY 416 CCAAGCTATGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 475

DB 361 CCAGCTTATGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 420  
QY 476 CCCACTGAGCTATGACCAACTCAATCTAGACAGGGGGTTACAACAGCCAGCTTAGA 535  
DB 421 CCCACTGAGCTATGACCAACTCAATCTAGACAGGGGGTTACAACAGCCAGCTTAGA 480  
QY 536 TATGACAGAGTAACTACAGTTATCCGAGGTACTGGAGCTAACCCATGACCGCTC 595  
DB 481 TATGACAGAGTAACTACAGTTATCCGAGGTACTGGAGCTAACCCATGACCGCTC 540  
QY 596 ACTGACCTTCACTCACTACCTCCAGAGCTATTCCTTCAACAGCCAGCTAGTTATGAT 655  
DB 541 ACTGACCTTCACTCACTACCTCCAGAGCTATTCCTTCAACAGCCAGCTAGTTATGAT 600  
QY 656 CAGAGCAGTTACTCTGACAGAAACCTATGGGCAACCGAGCAGTATGACAGCAGAGT 715  
DB 601 CAGAGCAGTTACTCTGACAGAAACCTATGGGCAACCGAGCAGTATGACAGCAGAGT 660  
QY 716 AGCTATGCTCAACAAAGCAGCTATGGGAGAGCTTCCCACTAGTTACCACTCAACT 775  
DB 661 AGCTATGCTCAACAAAGCAGCTATGGGAGAGCTTCCCACTAGTTACCACTCAACT 720  
QY 776 GGATCCACAGCCAGGCTCCAGTCAATATGACCAAGAGCAGCAGCTACGGGAGCAG 835  
DB 721 GGATCCACAGCCAGGCTCCAGTCAATATGACCAAGAGCAGCTACGGGAGCAG 780  
QY 836 AGTTCAATCCAGACAGACCAACCCAGTATGATGGTATGGGAGAGTCTGAGAGA 895  
DB 781 AGTTCAATCCAGACAGACCAACCCAGTATGATGGTATGGGAGAGTCTGAGAGA 840  
QY 896 TTTTCCGAGACAGAGAGAAACCGAGCATGATGGCTCTGATTAACCGGGGAGGGAGA 955  
DB 841 TTTTCCGAGACAGAGAGAAACCGAGCATGATGGCTCTGATTAACCGGGGAGGGAGA 900  
QY 956 GGGGAGTTTGTATGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1015  
DB 901 GGGGAGTTTGTATGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 960  
QY 1016 GGCAGGCTGAGAGAGGAGGAGGCTTCAATTAAGCTGTGAGCAACCTGATGAGAGCA 1075  
DB 961 GGCAGGCTGAGAGAGGAGGAGGCTTCAATTAAGCTGTGAGCAACCTGATGAGAGCA 1017  
QY 1076 GATCTTGAATCTAGGCGCCAGCTGTATGATCCAGATGAAGCTCTGACCAACGTCGAATTTAT 1135  
DB 1018 GATCTTGAATCTAGGCGCCAGCTGTATGATCCAGATGAAGCTCTGACCAACGTCGAATTTAT 1077  
QY 1136 GTTACAGAGATTAATGACAGTGTGACTCTAGATGATCTGCAACTCTTTAAGCAGTGT 1195  
DB 1078 GTTACAGAGATTAATGACAGTGTGACTCTAGATGATCTGCAACTCTTTAAGCAGTGT 1137  
QY 1196 GGGGTTGTTAAGTATGAACAAGAACTGGGCAACCATGATCCACTTACTGGAACAAG 1255  
DB 1138 GGGGTTGTTAAGTATGAACAAGAACTGGGCAACCATGATCCACTTACTGGAACAAG 1197  
QY 1256 GAAACAGAAAGCCCAAAGGCGATGCAAGTGTCTTATGAAGACCCACCACTGCAAG 1315  
DB 1198 GAAACAGAAAGCCCAAAGGCGATGCAAGTGTCTTATGAAGACCCACCACTGCAAG 1257  
QY 1316 GCTGCGGTGGAATGTTTATGAGGAGAAATTTTCAAGGAGAGCAACTTAAAGTCTCCCT 1375  
DB 1258 GCTGCGGTGGAATGTTTATGAGGAGAAATTTTCAAGGAGAGCAACTTAAAGTCTCCCT 1317  
QY 1376 GCTCGAAGAGGCTTCAATGAACAGTATGCGGGGTGTCTGCAACCCCGTGAAGGAGAG 1435  
DB 1318 GCTCGAAGAGGCTTCAATGAACAGTATGCGGGGTGTGTGCAACCCCGTGAAGGAGAG 1377  
QY 1436 GGCATGCCACCACTCTGAGAGTCAAGAGGCTCAAGAGGCTCTGGGGAGACCATG 1495  
DB 1378 GGCATGCCACCACTCTGAGAGTCAAGAGGCTCAAGAGGCTCTGGGGAGACCATG 1437  
QY 1496 GGTGCGATGGGAGGCGGTGAGAGATAGAGAGGCTTCCCTCAAGAGAGACCCCGGGGT 1555

Db 1438 GGTGCGATGGAGGAGCGGTGGAGAGATAGAGAGGCTTCCCTCCAGAGAGACCCCGAGG 1497  
 QY 1556 TCCCGAGGGAACCCCTCTGTGGAGAGAAAGTCCAGACCGAGCTGGAGCTGGCTGT 1615  
 Db 1498 TCCCGAGGGAACCCCTCTGTGGAGAGAAAGTCCAGACCGAGCTGGAGCTGGCTGT 1557  
 QY 1616 CCCAATCCGGGTTGGGAAACAGAACTTCGCTGGAGAGACAGAGTCAACAGTGTAA 1675  
 Db 1558 CCCAATCCGGGTTGGGAAACAGAACTTCGCTGGAGAGAGAGTCAACAGTGTAA 1617  
 QY 1676 GCCCAAGGCTGAAGGCTTCTCCCGCAACCTTTCGCGCCCGGGTGTGTGTGTG 1735  
 Db 1618 GCCCAAGGCTGAAGGCTTCTCCCGCAACCTTTCGCGCCCGGGTGTGTGTGTG 1677  
 QY 1736 AGAGTGGCCCTGTGGTGCATGCGGGAGAGAGAGTGTGCTCATGTGCTGTGTGCTCC 1795  
 Db 1678 AGAGTGGCCCTGTGGTGCATGCGGGAGAGAGAGTGTGCTCATGTGCTGTGTGCTCC 1737  
 QY 1796 GGTGGAATGTTCAAGGTTGGCGGTGGTGGAGACAGAGTGTGCTTCGCTGGTGGCGGGAGC 1855  
 Db 1738 GGTGGAATGTTCAAGGTTGGCGGTGGTGGAGACAGAGTGTGCTTCGCTGGTGGCGGGAGC 1797  
 QY 1856 ATGAGACCGAGTGGCTTGTGTGGAGAGACAGAGTGTGCTTCGCTGGTGGCGGGAGC 1915  
 Db 1798 ATGAGACCGAGTGGCTTGTGTGGAGAGACAGAGTGTGCTTCGCTGGTGGCGGGAGC 1857  
 QY 1916 TTGATGGAACAGATGGAG 1975  
 Db 1858 TTGATGGAACAGATGGAG 1917  
 QY 1976 GGGAGAGACCGTGCAG 2035  
 Db 1918 GGGAGAGACCGTGCAG 1977  
 QY 2036 GCATTGACTACAGATTATTTTAAACAGAAATGTTTAAATTTATTTATTTATTTAT 2095  
 Db 1978 GCATTGACTACAGATTATTTTAAACAGAAATGTTTAAATTTATTTATTTATTTAT 2037  
 QY 2096 TTATTAATTTGGTGCAGACATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 2155  
 Db 2038 TTATTAATTTGGTGCAGACATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 2097  
 QY 2156 TTGTGGAAGAAACATTTAAACAGTTAA 2184  
 Db 2098 TTGTGGAAGAAACATTTAAACAGTTAA 2126  
 RESULT 6  
 CR619493 2103 bp mRNA 11near HTC 21-JUL-2004  
 LOCUS full-length cDNA clone CS0D10761021 of Placenta Cot 25-normalized  
 DEFINITION  
 CR619493  
 ACCESSION  
 CR619493.1 GI:50500300  
 VERSION  
 HTC; CDSLT cDNA.  
 KEYWORDS  
 Homo sapiens (human)  
 SOURCE  
 Homo sapiens  
 ORGANISM  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE  
 AUTHORS  
 1 (bases 1 to 2103)  
 TITLE  
 Full-length cDNA libraries and normalization  
 JOURNAL  
 Unpublished  
 CONTACT  
 Contact: Feng Liang Email: fliang@lifestech.com URL:  
 http://fulllength.invitrogen.com/InvitrogenCorporation1600  
 Faraday Avenue  
 2 (bases 1 to 2103)  
 Genoscope.  
 REFERENCE  
 AUTHORS  
 Direct Submission  
 JOURNAL  
 Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage:  
 BP 191 91006 Evry cedex - FRANCE (E-mail: seque@genoscope.cns.fr  
 - Web: www.genoscope.cns.fr)  
 COMMENT  
 1st strand cDNA was primed with a NotI-oligo (dT) primer. Five prime

end enriched, double-strand cDNA was digested with Not I and cloned  
 into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library  
 was normalized. Library was constructed by Life Technologies, a  
 division of Invitrogen.  
 Location/Qualifiers  
 FEATURES  
 source  
 1..2103  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="CS0D10761021"  
 /tissue\_type="Placenta Cot 25-normalized"  
 /plasmid="pCMVSPORT\_6"  
 ORIGIN  
 Query Match 88.0%; Score 2103; DB 3; Length 2103;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 2103; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Db 56 GATTACAGTACTTATGACCACTGCGAGCGAGCGAGGCTTACAGTGTGCTTACACCGCCAG 115  
 1 GATTACAGTACTTATGACCACTGCGAGCGAGCGAGGCTTACAGTGTGCTTACACCGCCAG 60  
 QY 116 CCCACTCAAGATATGACAGACCAACCGAGCATATGGGCAACAAAGCTATGGAACCTAT 175  
 Db 61 CCCACTCAAGATATGACAGACCAACCGAGCATATGGGCAACAAAGCTATGGAACCTAT 120  
 QY 176 GGAAGCCCACTGATGCTATGACGCTATGACGCTGACAGACCACTGACACCTATGGGAGACC 235  
 Db 121 GGAAGCCCACTGATGCTATGACGCTATGACGCTGACAGACCACTGACACCTATGGGAGACC 180  
 QY 236 GCTATGACAACTTCTTATGAGAGAGCTCCCACTGTTATATCTCAACTGCCCCCAG 295  
 Db 181 GCTATGACAACTTCTTATGAGAGAGCTCCCACTGTTATATCTCAACTGCCCCCAG 240  
 QY 296 GCATACAGCAAGCCTGTCCAGGGGTATGCACTGTGCTTATGATACCACTGCTTAC 355  
 Db 241 GCATACAGCAAGCCTGTCCAGGGGTATGCACTGTGCTTATGATACCACTGCTTAC 300  
 QY 356 GTACACCAACCCAGGCTCTTATGAGGCTGATGAGCTGATATGAGCACTGAGCTGCTAT 415  
 Db 301 GTACACCAACCCAGGCTCTTATGAGGCTGATGAGCTGATATGAGCACTGAGCTGCTAT 360  
 QY 416 CCAGCTTATGGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 475  
 Db 361 CCAGCTTATGGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 420  
 QY 476 CCCACTGAGATGATCACTCACTCACTTATGACAGGGGTTTAAACAGGCCAGCTTATGAT 535  
 Db 421 CCCACTGAGATGATCACTCACTCACTTATGACAGGGGTTTAAACAGGCCAGCTTATGAT 480  
 QY 536 TATGACAGAGTACATCACTTATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 595  
 Db 481 TATGACAGAGTACATCACTTATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 540  
 QY 596 ACTGACCTTCATCTTACCTCTTACCTTACCTTACCTTACCTTACCTTACCTTACCTTACCT 655  
 Db 541 ACTGACCTTCATCTTACCTCTTACCTTACCTTACCTTACCTTACCTTACCTTACCTTACCT 600  
 QY 656 CAGAGCACTTCTCTGAGCAACACCTTATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 715  
 Db 601 CAGAGCACTTCTCTGAGCAACACCTTATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 660  
 QY 716 AGCTATGTCACAAAGGAGCTATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 775  
 Db 661 AGCTATGTCACAAAGGAGCTATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 720  
 QY 776 GATCTTACAGCCCAAGCTCAAGTCAATATAGCAACAGAGCAAGCAAGCAAGCAAGCAAG 835  
 Db 721 GATCTTACAGCCCAAGCTCAAGTCAATATAGCAACAGAGCAAGCAAGCAAGCAAGCAAG 780  
 QY 836 AGTTCAATTCGACAGAGCAACCCCACTAGATGAGGAGGAGGAGGAGGAGGAGGAGGAG 895  
 Db 781 AGTTCAATTCGACAGAGCAACCCCACTAGATGAGGAGGAGGAGGAGGAGGAGGAGGAG 840



QY 273 ATACTCTCCAACTGCCCCAGGCAATACAGCCAGGCTGTCCAGGGGTATGCGACTGTG 332  
 DB 181 ATACTCTCCAACTGCCCCAGGCAATACAGCCAGGCTGTCCAGGGGTATGCGACTGTG 240  
 QY 333 CTATGATACCACTGCTGCTACAGTCAACCAACCCAGGCTCTTATGAGCTCACTGTG 392  
 DB 241 CTATGATACCACTGCTGCTACAGTCAACCAACCCAGGCTCTTATGAGCTCACTGTG 300  
 QY 393 CATATGCACTGAGCTGCTTATCCAGCTTATGAGGAGGAGGAGGAGGAGGAGGAGG 452  
 DB 301 CATATGCACTGAGCTGCTTATCCAGCTTATGAGGAGGAGGAGGAGGAGGAGGAGG 360  
 QY 453 CAAGACCGAGAGTGAAGAACAGCCACTGAGCTAGTCAACTCAATCTAGGAGGAGG 512  
 DB 361 CAAGACCGAGAGTGAAGAACAGCCACTGAGCTAGTCAACTCAATCTAGGAGGAGG 420  
 QY 513 GTTACCAACGAGCCAGCTTATGAGTATGAGCAGAGTATCAAGTATATCCAGGAGT 572  
 DB 421 GTTACCAACGAGCCAGCTTATGAGTATGAGCAGAGTATCAAGTATATCCAGGAGT 480  
 QY 573 GAGGCTACCCGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 632  
 DB 481 GAGGCTACCCGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 540  
 QY 633 CTACACAGCCGAGTATGATATGAGCAGGATTAATCTGAGGAGGAGGAGGAGGAGG 692  
 DB 541 CTACACAGCCGAGTATGATATGAGCAGGATTAATCTGAGGAGGAGGAGGAGGAGG 600  
 QY 693 CGAGCAGCTATGAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 752  
 DB 601 CGAGCAGCTATGAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 660  
 QY 753 CCACTAGTTTCCCAACCCCAACTGAGTCTTACAGCCAGGAGGAGGAGGAGGAGG 812  
 DB 661 CCACTAGTTTCCCAACCCCAACTGAGTCTTACAGCCAGGAGGAGGAGGAGGAGG 720  
 QY 813 AGAGCAGCAGCTACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 872  
 DB 721 AGAGCAGCAGCTACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 780  
 QY 873 TTTATGAGGAGGAGTCTGAGGAGTTCGAGCAGGAGGAGGAGGAGGAGGAGGAGG 932  
 DB 781 TTTATGAGGAGGAGTCTGAGGAGTTCGAGCAGGAGGAGGAGGAGGAGGAGGAGG 840  
 QY 933 CTGATTAACCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 992  
 DB 841 CTGATTAACCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 900  
 QY 993 GGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1052  
 DB 901 GGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 960  
 QY 1053 GTGAGCCCATGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1112  
 DB 961 GTGAGCCCATGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1020  
 QY 1113 ACTCTGACCAAGTGTCAATTTATGATTAAGAGATTAATGACAGTGTATCTATGATG 1172  
 DB 1021 ACTCTGACCAAGTGTCAATTTATGATTAAGAGATTAATGACAGTGTATCTATGATG 1080  
 QY 1173 TGGCAGAGCTCTTATGAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1232  
 DB 1081 TGGCAGAGCTCTTATGAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1140  
 QY 1233 TGAATCCATCTTATGAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1292  
 DB 1141 TGAATCCATCTTATGAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1200  
 QY 1293 ATGAGAGCCCACTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1352  
 DB 1201 ATGAGAGCCCACTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1260  
 QY 1353 GAGAGCAAACTTAAAGTCTCCCTGTGCGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1412

DB 1261 GAGAGCAAACTTAAAGTCTCCCTGTGCGAGGAGGAGGAGGAGGAGGAGGAGG 1320  
 QY 1413 GTCTGACACCCGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1472  
 DB 1321 GTCTGACACCCGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1380  
 QY 1473 CAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1532  
 DB 1381 CAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1440  
 QY 1533 TCCCTCCAGAGAGACCCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1592  
 DB 1441 TCCCTCCAGAGAGACCCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1500  
 QY 1593 ACCGAGCTGAGAGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1652  
 DB 1501 ACCGAGCTGAGAGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1560  
 QY 1653 GAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1712  
 DB 1561 GAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1620  
 QY 1713 CGCCCCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1772  
 DB 1621 CGCCCCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1680  
 QY 1773 GCCTCATGAGATGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1832  
 DB 1681 GCCTCATGAGATGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1740  
 QY 1833 GTGGCTTCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1892  
 DB 1741 GTGGCTTCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1800  
 QY 1893 GGCCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1952  
 DB 1801 GGCCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1860  
 QY 1953 GAGGAGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2012  
 DB 1861 GAGGAGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1920  
 QY 2013 AGATGAGAGAGCCCGCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2072  
 DB 1921 AGATGAGAGAGCCCGCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1980  
 QY 2073 GTTTTAAATTTAATTTTCAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAA 2132  
 DB 1981 GTTTTAAATTTAATTTTCAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAA 2040  
 QY 2133 CTGTACTTTAGTATTTTTCACCAATTTGTGAAGAAACATTTAA 2175  
 DB 2041 CTGTACTTTAGTATTTTTCACCAATTTGTGAAGAAACATTTAA 2083

RESULT 8  
 CR594868  
 LOCUS full-length cDNA clone CS01064YE15 of Placenta Cot 25-normalized  
 DEFINITION  
 ACCESSION CR594868  
 VERSION CR594868.1 GI:50475675  
 KEYWORDS HTC; CDS; cDNA  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 2053)  
 AUTHORS Li, W.B., Gruber, C., Jessee, J., and Polyes, D.  
 TITLE Full-length cDNA libraries and normalization  
 JOURNAL Unpublished  
 REMARK Contact : Feng Liang Email : fliang@lifetech.com URL :









	Accession	Length	Species	Gene	Position	Strand	Sequence	Score
QY	1293	ATGAAGACCCACCCACCTGGCAAGGCGCGGTGAAA	1067	HTTGATGAGGAAAGATTTTCAAG	1355			
Db	1201	ATGAAGACCCACCCACCTGGCAAGGCGCGGTGAAA	1067	HTTGATGAGGAAAGATTTTCAAG	1260			
QY	1353	GGAGCAAACTTAAAGTCTTCTTGTGCGAAGACCTTCAATGAA	1321	CAGTATGCGGAGTG	1412			
Db	1261	GGAGCAAACTTAAAGTCTTCTTGTGCGAAGACCTTCAATGAA	1321	CAGTATGCGGAGTG	1320			
QY	1413	GTCTGCGCACCCCGTGAAGGCGAAGGCAATGCCACACAC	1321	CACTCCGTGGAAGTCCAGGAGGCC	1473			
Db	1321	GTCTGCGCACCCCGTGAAGGCGAAGGCAATGCCACACAC	1321	CACTCCGTGGAAGTCCAGGAGGCC	1380			
QY	1473	CAGGAGGTCTGTGGGGGACCCAGTGGGTGCAATGGGAGCGCGT	1321	GGAAGATGAGAGGCT	1532			
Db	1381	CAGGAGGTCTGTGGGGGACCCAGTGGGTGCAATGGGAGCGCGT	1321	GGAAGATGAGAGGCT	1440			
QY	1533	TCCCTTCGAAGAGAACCCCGGAGGTTCCCGAGGAGAACCCCT	1321	CTTGAAGAGAAACGTCACAGC	1592			
Db	1441	TCCCTTCGAAGAGAACCCCGGAGGTTCCCGAGGAGAACCCCT	1321	CTTGAAGAGAAACGTCACAGC	1500			
QY	1593	ACCGAGCTGAGAGACTGGCAGTGTCCCAATCCGGGTTGTGAAA	1321	ACCAAACTTTCGCTTGA	1652			
Db	1501	ACCGAGCTGAGAGACTGGCAGTGTCCCAATCCGGGTTGTGAAA	1321	ACCAAACTTTCGCTTGA	1566			
QY	1653	GAAAGAGTGCACCAAGTGTAAAGGCGCCAAAGCTGAAGGCTTCT	1321	CTCCCGGCACCCCTTTC	1712			
Db	1561	GAAAGAGTGCACCAAGTGTAAAGGCGCCAAAGCTGAAGGCTTCT	1321	CTCCCGGCACCCCTTTC	1620			
QY	1713	CGCCCCCGGAGTGTGTGTGTGCAAGGTGGCCCTGTGTGCGCAT	1321	GTGGCGGAGGAAAGAGGTG	1772			
Db	1621	CGCCCCCGGAGTGTGTGTGTGCAAGGTGGCCCTGTGTGCGCAT	1321	GTGGCGGAGGAAAGAGGTG	1680			
QY	1773	GCCTCATGAGATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT	1321	GT	1832			
Db	1681	GCCTCATGAGATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT	1321	GT	1740			
QY	1833	GTGGGTTTCCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT	1321	GT	1892			
Db	1741	GTGGGTTTCCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT	1321	GT	1800			
QY	1893	GCCCTGAGGAGGCGCCCTGTGTGTGTGTGTGTGTGTGTGTGTGT	1321	GT	1952			
Db	1801	GCCCTGAGGAGGCGCCCTGTGTGTGTGTGTGTGTGTGTGTGTGT	1321	GT	1860			
QY	1953	GAGGACCTGAGAAATGTGATTAAGGCGAGCACCGTCAGAGCGCAG	1321	AGATCGGCTTACT	2012			
Db	1861	GAGGACCTGAGAAATGTGATTAAGGCGAGCACCGTCAGAGCGCAG	1321	AGATCGGCTTACT	1920			
QY	2013	AGATGCAAGACCCCGCGAGAGGTGATTTGACTACCAAGATTTAT	1321	TTTAAACCAAGAAAT	2072			
Db	1921	AGATGCAAGACCCCGCGAGAGGTGATTTGACTACCAAGATTTAT	1321	TTTAAACCAAGAAAT	1980			
QY	2073	GTGTTAAATTTATATTCATATTTATATATGTGGCCCAACATTAT	1321	TAATCCTGT	2132			
Db	1981	GTGTTAAATTTATATTCATATTTATATATGTGGCCCAACATTAT	1321	TAATCCTGT	2040			
QY	2133	CTGTACTTGTAGATTTTTCACCATTTGTGAAGA	2165					
Db	2041	CTGTACTTGTAGATTTTTCACCATTTGTGAAGA	2073					
RESULT 10								
CR602561								
LOCUS	full-length cDNA clone CS001068	2070 bp	nrna	linear	HTC 21-JUN-2004			
DEFINITION	Of Homo sapiens (human).				Cot 25-normalized			
ACCESSION	CR602561							
VERSION	CR602561.1							
KEYWORDS	HTC; cDNA.							
SOURCE	Homo sapiens (human)							
ORGANISM	Homo sapiens							
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;							
	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.							

REFERENCE	AUTHORS	TITLE	JOURNAL	REMARK
1	(bases 1 to 2070)	Li, W.B., Gruber, C., Jessee, J. and Polayes, D.	Full-length cDNA libraries and normalization	Unpublished
2	(bases 1 to 2070)	Paradey Avenue	Invitrogen Corporation	1600
3	(bases 1 to 2070)	Genoscope	Genoscope	
4	Submitted (20-JUN-2004)	FRANCE (E-mail: sequef@genoscope.cns.fr)	BP 191 91006 EVRY cedex	
5	- web : www.genoscope.cns.fr			
6	1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime			
7	end enriched, double-strand cDNA was digested with NotI and cloned			
8	into the NotI and EcoR V sites of the pCMVSPORT 6 vector. Library			
9	was normalized. Library was constructed by Life Technologies, a			
10	division of Invitrogen			
11	Location/Qualifiers			
12	1. .2070			
13	/organism="Homo sapiens"			
14	/mol_type="mRNA"			
15	/db_xref="taxon:9606"			
16	/clone="CS0D1068YK23"			
17	/tissue_type="Placenta Cot 25-normalized"			
18	/plasmid="pCMVSPORT_6"			
19	ORIGIN			
20	Query Match	86.1%	Score 2057;	DB 3; Length 2070;
21	Best Local Similarity	99.9%	Pred. No. 0;	
22	Matches 2070; Conservative	0;	Mismatches 0;	Indels 3; Gaps 1;
23	56 GATTACAGTACTTATAGCCAGCTGACAGCGAGGCTACAGTGTCTTACACCGCCAG	115		
24	1 GATTACAGTACTTATAGCCAGCTGACAGCGAGGCTACAGTGTCTTACACCGCCAG	60		
25	116 CCCACTCAGAGATATGACAGACCAACCGCATATGGGCAACAAAGCTATGAACTTAT	175		
26	61 CCCACTCAGAGATATGACAGACCAACCGCATATGGGCAACAAAGCTATGAACTTAT	120		
27	176 GGACAGCCCATGATGTCAGTATATCCAGGCTCAGACCACTGCAACCTATATGGAGACC	235		
28	121 GGACAGCCCATGATGTCAGTATATCCAGGCTCAGACCACTGCAACCTATATGGAGACC	180		
29	236 GCCTATGCACTTCTTATGACAGCGCTCCACTGGTATATCTACTCACAAGCCCCAG	295		
30	181 GCCTATGCACTTCTTATGACAGCGCTCCACTGGTATATCTACTCACAAGCCCCAG	240		
31	296 GCATACAGCCAGCGCTGTCAGAGGATATGCACTGGTCTTATGATACCACTGTCTACA	355		
32	241 GCATACAGCCAGCGCTGTCAGAGGATATGCACTGGTCTTATGATACCACTGTCTACA	300		
33	356 GTACACCAACCCAGGCTCTTATGACAGCTAGTCTGATATGGCACTCAGCTGCTTAT	415		
34	301 GTACACCAACCCAGGCTCTTATGACAGCTAGTCTGATATGGCACTCAGCTGCTTAT	360		
35	416 CCAGCCTATGGGCGAGCGAGCGACCTGCACTTCAAGACCGAGAGATGGAACAAG	475		
36	361 CCAGCCTATGGGCGAGCGAGCGACCTGCACTTCAAGACCGAGAGATGGAACAAG	420		
37	476 CCAGCCTATGGGCGAGCGAGCGACCTGCACTTCAAGACCGAGAGATGGAACAAG	535		
38	421 CCAGCCTATGGGCGAGCGAGCGACCTGCACTTCAAGACCGAGAGATGGAACAAG	480		
39	536 TATGACAGAGTAACTACAGTTATCCAGGTAACCTGGAGCTACCCATGACCGCATGC	595		
40	481 TATGACAGAGTAACTACAGTTATCCAGGTAACCTGGAGCTACCCATGACCGCATGC	540		
41	596 ACTGACCTCATCTTACCTCTTACAGCACTTATCTTACAGCGGAGCTAGTTATGAT	655		
42	541 ACTGACCTCATCTTACCTCTTACAGCACTTATCTTACAGCGGAGCTAGTTATGAT	600		
43	656 CAGGACAGTAACTCTCAGCAACCACTATGGGCAACCGAGACCTATGACAGCAGAGT	715		

Db 601 CAGAGCACTTACTCTGACGAGAACCTATGGGCAACCGACGCTATGAGCAGCAGAGT 660  
 Qy 716 AGCTATGTCACAAAGACGCTATGGGACAGCCTTCCACTGTTACCACTCCAACT 775  
 Db 661 AGCTATGTCACAAAGACGCTATGGGACAGCCTTCCACTGTTACCACTCCAACT 720  
 Qy 776 GATTCCTACGACGAGTCCAGTCAATATAGCCACAGAGCAGCAGTCCGGGACGAG 835  
 Db 721 GATTCCTACGACGAGTCCAGTCAATATAGCCACAGAGCAGCAGTCCGGGACGAG 780  
 Qy 836 AGTTCAATCCGACAGCAGCAGCAGCAGTGGGTGTTATGGGACAGGAGTCCGAGAG 895  
 Db 781 AGTTCAATCCGACAGCAGCAGCAGCAGTGGGTGTTATGGGACAGGAGTCCGAGAG 840  
 Qy 896 TTTTCCGACAGCAGAGAAACCGAGCAGTATGTCCTGATTAACCGGGACAGGAGAG 955  
 Db 841 TTTTCCGACAGCAGAGAAACCGAGCAGTATGTCCTGATTAACCGGGACAGGAGAG 900  
 Qy 956 GGGGAGTTTGAATGTCGAGGAGCAGAGAGTGGGGGAGGAGAGGAGGAGGAGTGGAGT 1015  
 Db 901 GGGGAGTTTGAATGTCGAGGAGCAGAGAGTGGGGGAGGAGAGGAGGAGGAGTGGAGT 960  
 Qy 1016 GGCAGCGCTGAGAGAGCAGGAGTGGCTTCAATAGCTGAGAGCAGCAGTGAAGAGCA 1075  
 Db 961 G---GCGCTGAGAGAGCAGGAGTGGCTTCAATAGCTGAGAGCAGCAGTGAAGAGCA 1017  
 Qy 1076 GATCTTGAATGAGGCCCACTGATGATCCAGATGAGAGCAGTCTGACCAAGTGCATTTAT 1135  
 Db 1018 GATCTTGAATGAGGCCCACTGATGATCCAGATGAGAGCAGTCTGACCAAGTGCATTTAT 1077  
 Qy 1136 GTACCAAGATTTAATGACAGTGAATCTGATGATGAGAGCAGTCTTAAAGCAGTGT 1195  
 Db 1078 GTACCAAGATTTAATGACAGTGAATCTGATGATGAGAGCAGTCTTAAAGCAGTGT 1137  
 Qy 1196 GGGGTGTTAAGATGACCAAGAGAGCTGGGCAACCCATGATCCACATCTGAGCAG 1255  
 Db 1138 GGGGTGTTAAGATGACCAAGAGAGCTGGGCAACCCATGATCCACATCTGAGCAG 1197  
 Qy 1256 GAAACAGAGAAACCCAAAGGCGATGCCAGTGTCTTATGAGAGCCCACTGCCAG 1315  
 Db 1198 GAAACAGAGAAACCCAAAGGCGATGCCAGTGTCTTATGAGAGCCCACTGCCAG 1257  
 Qy 1316 GCTGCGGTGAGATGTTGATGGAGAGATTTTCAAGGAGCAACTTAAAGTCTCCCT 1375  
 Db 1258 GCTGCGGTGAGATGTTGATGGAGAGATTTTCAAGGAGCAACTTAAAGTCTCCCT 1317  
 Qy 1376 GCTCGAGAGAGCCTCCATGACATGATGCGGGTGTCTGCCACCCCGTGAAGGAGCA 1435  
 Db 1318 GCTCGAGAGAGCCTCCATGACATGATGCGGGTGTCTGCCACCCCGTGAAGGAGCA 1377  
 Qy 1436 GGCATGCGACACCACTCCGTGAGGAGTCCAGAGAGCCCAAGAGTCTGGGGAGCCATG 1495  
 Db 1378 GGCATGCGACACCACTCCGTGAGGAGTCCAGAGAGCCCAAGAGTCTGGGGAGCCATG 1437  
 Qy 1496 GGTGCGATGGAGGCGGTGAGGAGATGAGAGAGCTTCCCTCCAGAGGAGCCCGGGGT 1555  
 Db 1438 GGTGCGATGGAGGCGGTGAGGAGATGAGAGAGCTTCCCTCCAGAGGAGCCCGGGGT 1497  
 Qy 1556 TCCCGAGAGAGCCTCTGAGAGAGAGAGCCTCCAGACCGAGCTGAGAGCTGGCAGTGT 1615  
 Db 1498 TCCCGAGAGAGCCTCTGAGAGAGAGAGCCTCCAGACCGAGCTGAGAGCTGGCAGTGT 1557  
 Qy 1616 CCCATTCGGGTGTGAGAGAGAGAGCTTGGCTGAGAGAGAGTGCACCACTGTATAG 1675  
 Db 1558 CCCATTCGGGTGTGAGAGAGAGAGCTTGGCTGAGAGAGAGTGCACCACTGTATAG 1617  
 Qy 1676 GCGCCAAAGCCTGAGAGCTTCTCCGCGCACCTTCCGCGCGCGGGGTGATGCTGGG 1735  
 Db 1618 GCGCCAAAGCCTGAGAGCTTCTCCGCGCACCTTCCGCGCGCGGGGTGATGCTGGG 1677  
 Qy 1736 AGAGTGGCCCTGTGAGAGTGGGAGAGAGAGTGGCTCATGATGCTGTGTCTCC 1795

Db 1678 AGAGTGGCCCTGTGAGCATGCGGGAGAGAGAGTGGCTCATGATGCTGTGTCC 1737  
 Qy 1796 GGTGAGATTTGAGAGTGGCCGTGTGAGAGAGAGTGGCTTCCGTGGTGGCGGGG 1855  
 Db 1738 GGTGAGATTTGAGAGTGGCCGTGTGAGAGAGAGTGGCTTCCGTGGTGGCGGGG 1797  
 Qy 1856 ATGACCCGAGTGGCTTGTGAGAGAGAGAGTGGCTTGTGGGGGCGCCCTGAGACT 1915  
 Db 1798 ATGACCCGAGTGGCTTGTGAGAGAGAGAGTGGCTTGTGGGGGCGCCCTGAGACT 1857  
 Qy 1916 TTGATGAGAACAGATGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1975  
 Db 1858 TTGATGAGAACAGATGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1917  
 Qy 1976 GGGAGCAGCCTGAG 2035  
 Db 1918 GGGAGCAGCCTGAG 1977  
 Qy 2036 GCATTGACTACAGATTTATTTTAAACAGAGAAATGTTTAAATTTATTTATTTAT 2095  
 Db 1978 GCATTGACTACAGATTTATTTTAAACAGAGAAATGTTTAAATTTATTTATTTAT 2037  
 Qy 2096 TTATATGTTGGCCACACATTTATTTATTTCC 2128  
 Db 2038 TTATATGTTGGCCACACATTTATTTATTTCC 2070  
 RESULT 11  
 AK034755 2373 bp mRNA 1linear HTC 03-APR-2004  
 LOCUS AK034755  
 DEFINITION Mus musculus 12 days embryo embryonic body between diaphragm region and neck cDNA, RIKEN full-length enriched library, clone:9430031M11  
 ACCESSION AK034755 GI:26084174  
 VERSION AK034755.1  
 KEYWORDS HTC; CAP trapper.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 REFERENCE  
 1 Carninci, P. and Hayashizaki, Y.  
 High-efficiency full-length cDNA cloning  
 Meth. Enzymol. 303, 19-44 (1999)  
 2 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.  
 Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes  
 Genome Res. 10 (10), 1617-1630 (2000)  
 TITLE  
 JOURNAL MEDLINE 2049374  
 PUBMED 11042159  
 AUTHORS  
 3 Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Kono, H., Akiyama, J., Nishi, K., Katsunai, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Harada, M., Nishino, T., Harada, A., Yamamoto, K., Watanabe, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujisawa, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Wataniki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsunaga, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.  
 RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer  
 Genome Res. 10 (11), 1757-1771 (2000)  
 JOURNAL MEDLINE 20530913  
 PUBMED 11076861  
 REFERENCE  
 4 The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.  
 Functional annotation of a full-length mouse cDNA collection  
 Nature 409, 685-690 (2001)  
 TITLE  
 JOURNAL MEDLINE  
 REFERENCE 5



Db 1261 AAAGCTTAAGGAGCGCAAGTGTCTATGAAGTCCAACTGAAAGGCTCCGT 1320

Qy 1324 GGAATGTTGATGGAAAGATTTTCAAGGAGCAAACTTAAAGTCTCCCTGCGAA 1383

Db 1321 GGAATGTTGATGGAAAGATTTTCAAGGAGCAAACTTAAAGTCTCTGCGGAA 1380

Qy 1384 GAAGCTCCAAATGAAGATATGCGGGGTGTCTGCCACCCCGGAGGGGAGAGCGATGCC 1443

Db 1381 GAAGCTCCAAATGAAGATATGCGGGAGGAGATCCACTGTGAAGGAGGAGGAGATGCC 1440

Qy 1444 ACCACCACTCCGTGAGATCCAGAGAGCCAGAGAGTCTCTGGAGGAGCCATGGGTGCAT 1503

Db 1441 ACCACCACTCCGTGAGATCCAGAGAGCCAGAGAGTCTCTGGAGGAGCCATGGGTGCAT 1500

Qy 1504 GGGAGGCGGTGAGAGATGAGAGAGGCTTCCCTCCAAAGAGACCCCGGGGTTCCTCCGAG 1563

Db 1501 GGGAGGCGGTGAGAGAGAGAGAGGAGGCTTCCCTCCAAAGAGGAGGAGGAGGAGGAG 1560

Qy 1564 GAACCCCTCTGAGAGAGAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1623

Db 1561 GAACCCCTCTGAGAGAGAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1620

Qy 1624 GGGTGTGAGAAACAGAACTTCCCTGAGAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAG 1683

Db 1621 GGGCTGTGAGAAACAGAACTTCCCTGAGAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAG 1680

Qy 1684 GCGTGAAGGCTTCTCCCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1743

Db 1681 GCGGAGGAGGCTTCTCCCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1740

Qy 1744 CCGTGTGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1803

Db 1741 CCGTGTGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1800

Qy 1804 GTTCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1863

Db 1801 GTTCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1860

Qy 1864 AGGTGCTTGTGAGAGAGAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1923

Db 1861 AGGTGCTTGTGAGAGAGAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1920

Qy 1924 ACAGATGAGAGAGAGAGAGAGAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1983

Db 1921 ACAGATGAGAGAGAGAGAGAGAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1980

Qy 1984 CCGTCAAGAGGAGGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2043

Db 1981 CCGTCAAGAGGAGGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2033

Qy 2044 TACCAAGATTTATTTTAAACGAGAAATGTTTAAATTTAAATTTTAAATTTTAAATTTTAAATTT 2103

Db 2034 GACACAGATTTATTTTAAACGAGAAATGTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTT 2093

Qy 2104 TTGGCCCAACATTAATTAATTTCTGTCGATTAATTTTAAATTTTAAATTTTAAATTTTAAATTTT 2163

Db 2094 TTGGCCCAACATTAATTAATTTCTGTCGATTAATTTTAAATTTTAAATTTTAAATTTTAAATTTT 2152

Qy 2164 GAACATTTAAACAGTTAAATGATGTCGAGAGTTT----- 2204

Db 2153 GAACATTTAAACAGTTAAATGATGTCGAGAGTTT----- 2212

Qy 2205 -----TTTCTCTCTCTTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTT 2255

Db 2213 GATTTGTTCTTTTCTTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTT 2272

Qy 2256 CTTTGAGAGATGTCGATTAATTTGAGAAACAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2315

Db 2273 CTTTGAGAGATGTCGATTAATTTGAGAAACAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2331

Qy 2316 GTTCATGTTGTGATGTTT----- 2340

Db 2332 GTTCATGTTGTGATGTTT----- 2356

## RESULT 12

AK019460

## LOCUS

AK019460

## DEFINITION

AK019460

## ACCESSION

AK019460

## VERSION

AK019460.1

## KEYWORDS

Mus musculus (house mouse)

## SOURCE

Mus musculus

## ORGANISM

Mus musculus

## REFERENCE

Carninci, P. and Hayashizaki, Y.

## AUTHORS

Carninci, P. and Hayashizaki, Y.

## TITLE

High-efficiency full-length cDNA cloning

## JOURNAL

Meth. Enzymol. 303, 19-44 (1999)

## MEDLINE

99279252

## PUBMED

10349636

## REFERENCE

Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,

## AUTHORS

Itoh, M., Komori, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.

## TITLE

Normalization and subtraction of cap-trapper-selected cDNAs to

## JOURNAL

Genome Res. 10 (10), 1617-1630 (2000)

## MEDLINE

20493374

## PUBMED

11042159

## REFERENCE

Shibata, K., Itoh, M., Aizawa, K., Nagaoaka, S., Sasaki, N., Carninci, P.,

## AUTHORS

Komori, H., Akiyama, J., Nishi, K., Katsunuma, T., Tashiro, H., Itoh, M.,

## TITLE

Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Aishine, T., Harada, A.,

## JOURNAL

Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,

## MEDLINE

Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watanabe, M.,

## PUBMED

Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsunaga, S., Kawai, J.,

## REFERENCE

Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A., and Hayashizaki, Y.

## AUTHORS

RIKEN integrated sequence analysis (RISA) system-384-format

## JOURNAL

Genome Res. 10 (11), 1757-1771 (2000)

AK019460 2107 bp mRNA linear HTC 03-APR-2004

Mus musculus 18 days pregnant adult female placenta and extra

embryonic tissue cDNA, RIKEN full-length enriched library,

clone:38304117B11 product:EWing sarcoma homolog, full insert

sequence.

AK019460

AK019460.1 GI:12859676

HTC, CAP trapper.

Mus musculus (house mouse)

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1

Carninci, P. and Hayashizaki, Y.

High-efficiency full-length cDNA cloning

Normalization and subtraction of cap-trapper-selected cDNAs to

prepare full-length cDNA libraries for rapid discovery of new genes

Genome Res. 10 (10), 1617-1630 (2000)

2

Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,

Itoh, M., Komori, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.

3

Shibata, K., Itoh, M., Aizawa, K., Nagaoaka, S., Sasaki, N., Carninci, P.,

Komori, H., Akiyama, J., Nishi, K., Katsunuma, T., Tashiro, H., Itoh, M.,

Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Aishine, T., Harada, A.,

Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,

Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watanabe, M.,

Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsunaga, S., Kawai, J.,

Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A., and Hayashizaki, Y.

4

RIKEN integrated sequence analysis (RISA) system-384-format

sequencing pipeline with 384 multicapillary sequencer

Genome Res. 10 (11), 1757-1771 (2000)

5

The RIKEN Genome Exploration Research Group Phase II Team and the

PANTOM Consortium.

Functional annotation of a full-length mouse cDNA collection

Nature 409, 685-690 (2001)

6

The PANTOM Consortium and the RIKEN Genome Exploration Research

Group Phase I & II Team.

Analysis of the mouse transcriptome based on functional annotation

of 60,770 full-length cDNAs

Nature 420, 563-573 (2002)

7

Adachi, J., Aizawa, K., Akahira, S., Akimura, T., Arai, A., Aono, H.,

Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Fukunishi, Y.,

Furuno, M., Hanagaki, T., Hara, A., Hayatsu, N., Hiramoto, K.,

Hirooka, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Izawa, M.,

Kasukawa, T., Kato, H., Kawai, J., Kojima, Y., Komori, H., Kouda, M.,

Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Nishi, K.,

Nomura, K., Numazaki, R., Ohno, M., Okazaki, Y., Okido, T., Owa, C.,

Saito, H., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki, D.,

Shibata, K., Shibata, Y., Shingawa, A., Shiraki, T., Sogabe, Y.,

Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F., Tanaka, T.,

Tejima, Y., Toyota, T., Yamamura, T., Yasunishi, A., Yoshida, K.,

Yoshino, M., Yamamoto, M., and Hayashizaki, Y.

8

Submitted (18-AUG-2000) Yoshihide Hayashizaki, The Institute of

Physical and Chemical Research (RIKEN), Laboratory for Genome

Exploration Research Group, RIKEN Genomic Sciences Center (GSC),

RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Teikyo-ku, Yokohama,

Kanagawa 230-0045, Japan (E-mail: genome-res@sc.riken.jp,

9





Db	Accession	Source	Length	Reference
Db	1318	TCAGGAAGCAAATTAAAGTGTCTTTGCCCCGAAAGAGGCTCCATATGAACACATATCGC	1377	
Qy	1408	GGGTGCTTCGCGACCCCGTGAAGGCGAAGCATGCCACACACTCCGTGAGGTCCAGG	1467	
Db	1378	GGGAGGCAATGCGCACTCGTAGAGGCGAGGGGAGTGCACACACACTTCGTGAGGTCTCGG	1437	
Qy	1468	AGGCGCCAGAGAGTCTGGGGGAGCCCATAGGGTTCGATGGGAAGCCGTGGAGAGATTAAGG	1527	
Db	1438	TGGGCCCAAGAGAGCCCTGAGGAGCCCATAGGGTTCGATGGGAAGCCGTGGAGAGACAAGG	1497	
Qy	1528	AGGCTTCCCTCCAGAGAGGACCCCGGGGTTCCGAGAGGAAACCCCTCTGGAGAGAGAAAGT	1587	
Db	1498	GGGCTTCCCTCCAGAGAGGACCCCGAGAGCTCCAGAGAGAAACCCCTCTGGAGAGAGAAATGT	1557	
Qy	1588	CCAGCAACCGAGCTGAGAGCTGGCAGTGTCCCAATCCGGGTTGTGAAAACAGAACTTTCG	1647	
Db	1558	CCAGCAACCGAGCTGAGAGCTGGCAGTGTCCCAATCCGGGCTGTGAAAACAGAACTTTCG	1617	
Qy	1648	CTGAGAGAACAGAGTGCACACAGATGAAGAGCCCAAGAGCTGAAGCTTCCCTCCGCGAC	1707	
Db	1618	TTGGAGAAACAATAATGCAACAGATGAAGAGCCCTGAAGCGAGGCTTCTCTCCGCGAC	1677	
Qy	1708	CTTTCGCGCCCCGGGTGTGATTCGTCGAGAGTGGCCCTGTGTGACATGCGGGAGAG	1767	
Db	1678	CTTTCGCGCCCCGGGTGTGATTCGTCGAGAGTGGCCCTGTGTGACATGCGGGAGAG	1737	
Qy	1768	AGGTGGCCTCATGAGATCGTGTGTGTCCCGATGGAATGTTCAAGAGTGGCCGTGTGAG	1827	
Db	1738	AGGAGGACTCATATGAGACGTGGTGGTCTTGAGAGAAATGTTCAAGAGTGGCAAGGTGAG	1797	
Qy	1828	CAGAGTGTGCTTTCGTGTGTGTCGCGGGCATGAGCCAGATGTGCTTGTGTGAGAGAACG	1887	
Db	1798	CAGAGGAGGCTTTCGAGGTGTGTCGCGGGCATGAGCCAGATGTGCTTGTGTGAGAGAACG	1857	
Qy	1888	AGGTGGCCTGTGGGGGGCCCGCTGAGACTTTGATGGAACAGATGGGAGAGAGAGAGG	1947	
Db	1858	AGGTGTGTCTGGGGGGCCCTCTGAGACTTTGATGGAACAGATGGGAGAGAGAGGCGG	1917	
Qy	1948	ACGTGAGAGACCTGGAATAATGATTAAGGCGAGACCGTCAAGAGCCGAGAGATTCGGCC	2007	
Db	1918	ACGTGAGAGACCTGGAATAATGATTAAGGCGAGACCGTCAAGAGCCGAGAGATTCGGCC	1977	
Qy	2008	CTACTAGATGCAAGAGACCCCGAGAGCTGCATGCTACGAGATTATTTTAAACAG	2067	
Db	1978	CTACTAGA-----GACCTGACAGAGCTGCATGCGACGAGATTATTTTAAACAG	2030	
Qy	2068	AAAATGTTTAAATTATTAATTCATATTATTAATGTTGGGCAACAATATGATTATTC	2127	
Db	2031	AAAATGTTTAAATTATTAATTCATATTATTAATGTTGGGCAACAATATGATTATTC	2090	
Qy	2128	CTGTCTGTACTTTAG 2143		
Db	2091	CTGTCTGTACTTTAG 2106		
RESULT 13	AK049743	2269 bp	mRNA	linear
LOCUS	AK049743			
DEFINITION	Mus musculus 12 days embryo spinal cord cDNA, RIKEN full-length			
	enriched library, clone: C50046A18			
	product: Bwing sarcoma homolog,			
	full insert sequence.			
ACCESSION	AK049743			
VERSION	AK049743.1	GI:26093614		
KEYWORDS	HTC, CAP trapper.			
SOURCE	Mus musculus (house mouse)			
ORGANISM	Mus musculus			
	Eutheria; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
REFERENCE	1	Carninci, P. and Hayashizaki, Y.		
AUTHORS		High-efficiency full-length cDNA cloning		
TITLE				
JOURNAL		Meth. Enzymol. 303, 19-44 (1999)		

JOURNAL MEDLINE  
PUBMED 99279253  
REFERENCE 10349636

AUTHORS Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Komno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.

TITLE Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes

JOURNAL MEDLINE  
PUBMED 20499374  
REFERENCE 11042159

AUTHORS 3 Shibata, K., Itoh, M., Aizawa, K., Nagaoaka, S., Sasaki, N., Carninci, P., Komno, H., Akiyama, J., Nishi, K., Katsunai, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Harada, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujikake, S., Inoue, K., Togawa, Y., Izawa, M., Onate, S., Watabiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsunaga, S., Kawai, D., Okazaki, Y., Muramatsu, M., Inoue, Y., Kita, A. and Hayashizaki, Y.

TITLE RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multichannel sequencer

JOURNAL MEDLINE  
PUBMED 20530913  
REFERENCE 11076861

AUTHORS 4 Genome Res. 10 (11), 1757-1771 (2000)

TITLE THE RIKEN Genome Exploration Research Group Phase II Team and the PANTOM Consortium.

JOURNAL MEDLINE  
PUBMED Nature 409, 685-690 (2001)  
REFERENCE 11076861

AUTHORS 5 The PANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.

TITLE Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs

JOURNAL MEDLINE  
PUBMED Nature 420, 563-573 (2002)  
REFERENCE 11076861

AUTHORS 6 (bases 1 to 2269)

TITLE Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hirooka, T., Hirose, T., Horii, F., Imctani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Komno, H., Koude, M., Koya, S., Kuhihara, C., Macusayama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numasaki, R., Ohno, M., Ohsato, N., Okazaki, Y., Saito, R., Saitoh, H., Sekai, C., Sekai, K., Sakazume, N., Sano, H., Saeki, D., Shibata, K., Shingawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takai-Akahira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.

TITLE Direct Submission

JOURNAL MEDLINE  
PUBMED Submitted (16-JUL-2001)  
REFERENCE 11076861

AUTHORS Yoshinori Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.jp).

TITLE URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216

COMMENT cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.

FEATURES Please visit our web site for further details.  
SOURCE URL: http://genome.gsc.riken.jp/  
location/Qualifiers  
1..2269  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/script="CS7B1/60"  
/db\_xref="PANTOM\_DB:CS30046A18"  
/clone="CS30046A18"  
/c1one\_lib="RIKEN full-length enriched mouse cDNA library"







Query Match 45.8%; Score 1094.4; DB 3; Length 1379;  
 Best Local Similarity 89.7%; Pred. No. 1e-247;  
 Matches 1249; Conservative 0; Mismatches 96; Indels 47; Gaps 5;

QY 1011 GAATGGGACGCGCTGAGAGCGAGGTGGCTTCAATAGCTGAGTGAAGCCATGATAG 1070  
 DB 1 GACTGGGACGCGCTGAGAGCGAGGTGGCTTCAATAGCTGAGTGAAGCCATGATAG 60

QY 1071 GACCAAGTCTTGAATCTAGGCGCACTGTAGATCCAGATGAGAACTCTGACAAAGTGC 1130  
 DB 61 GACCAAGTCTTGAATCTAGGCGCTTCTATGATCCAGATGAGAACTCTGACAAAGTGC 120

QY 1131 TTTATGACAAAGATTTAAATGACAGTGTGACTGATGATCTGAGCACTTTTAAAC 1190  
 DB 121 TTTATGACAAAGATTTAAATGACAAATGTGACTGATGATCTGAGCACTTTTAAAC 180

QY 1191 AGTGTGGGCTTGTAAATGACAAAGAACTGGGCAACCATGATCAGATCTTACCTG 1250  
 DB 181 AGTGTGGGCTTGTCAAGATGACAAAGAACTGGGCAACCATGATCAGATCTTACCTG 240

QY 1251 ACAAGAAACAGAAAGCCCAAGGCGATGCCAGTGTCTATGAAAGCCCACTG 1310  
 DB 241 ATAAAGAGACAGAAAGCTTAAAGGCGACGCAAGTGTCTATGAAAGCCCACTG 300

QY 1311 CCAAGGCTGCGGTGAAATGCTTGAATGAGAAATTTTCAAGGAGCACTTAAATCT 1370  
 DB 301 CCAAGGCTGCGGTGAAATGCTTGAATGAGAAATTTTCAAGGAGCACTTAAATCT 360

QY 1371 CCCCTGCTCGGAGAAAGCTCCAAATGAAAGTATGCGGGGTGGCTGCGCAACCCCGT 1430  
 DB 361 CTCCTGCTCGGAGAAAGCTCCCAATGAAAGTATGCGGGGTGGCTGCGCAACCCCGT 420

QY 1431 GCAGAGCATGCCACCACTCCGCTGAGAGTCCAGAGGCGCCAGAGTCTTGGGAGAC 1490  
 DB 421 GCAGAGGAGTGCACCACTCCGCTGAGAGTCTTGGGAGGCGCCAGAGGCGCTGAGAGAC 480

QY 1491 CCAATGGTCCGATGAGAGGCGCTGAGAGATGAGAGGCTTCCCTCAAGAGACCC 1550  
 DB 481 CCAATGGTCCGATGAGAGGCGCTGAGAGATGAGAGGCTTCCCTCAAGAGGCGCC 540

QY 1551 GGGGTTCCCGAGGAAACCCCTCTGAGAGAGAAAGTCCAGCAACCGAGTGGAGACTGC 1610  
 DB 541 GAGGCTCTCAAGAGAAACCCCTCTGAGAGAGAAAGTCCAGCAACCGAGTGGAGACTGC 600

QY 1611 AGTGTCCCAATCCGAGTGTGAGAAACCAATCTTCCCTGAGAGAAAGATGCAACAGT 1670  
 DB 601 AGTGTCCCAATCCGAGGCTGTGAGAAACCAATCTTCCCTGAGAGAAAGATGCAACAGT 660

QY 1671 GTAAAGCCCAAAAGCTGAAAGCTTCTCCGCGCAACCTTTCCGCGCGGAGTGTATC 1730  
 DB 661 GTAAAGCCCTTAAGCCCGAGAGGCTTCTCCGCGCAACCTTTCCGCGGAGTGTATC 720

QY 1731 GTAGCAGAGGTGGCGCTGTGAGCAATGCGGAGAGAGAGGTGCTCATGATCTGTGTG 1790  
 DB 721 GTAGCAGAGGTGGCGCTGTGAGCAATGCGGAGAGAGAGGTGCTCATGATCTGTGTG 780

QY 1791 GTCCCGTGAATGTTTCAAGTGTGCGCTGTGAGCAAGAGTGTGCTTCCGTGTGTGCC 1850  
 DB 781 GTCTGTGAGAGAAATTTTCAAGTGTGCGAGAGAGAGAGAGAGAGAGAGAGAGTGTG 840

QY 1851 GGGGACATGACCGAGGTGCTTGTGTGAGAGAAAGAGAGTGTGCGGAGGCGCTG 1910  
 DB 841 GTGGAATGACCGAGGTGCTTGTGTGAGAGAAAGAGAGTGTGCGGAGGCGCTG 900

QY 1911 GACCTTTGATGAGACAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1970  
 DB 901 GACCTTTGATGAGACAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 960

QY 1971 ATAAAGCGAGCAACCGTCAAGAGCGAGAGATGCGGCTCTATGATGAGAGAGAGAG 2030  
 DB 961 ATAAAGCGAGCAACCGTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1013

QY 2031 GAGCTGATGACTACAGATTTATTTTAAACCAAGAAATGTTTAAATTAATATTC 2090  
 DB 1014 GAGCTGATGAGACAGACAGATTTATTTTAAACCAAGAAATGTTTAAATTAATATTC 1073

QY 2091 CATATTTAATATGTCGCCCAACATATGATATTTCTGTGCTGATCTTATGATTTT 2150  
 DB 1074 CATATTTAATATGTCGCCCAACATATGATATTTCTGTGCTGATCTTATGATTTT 1132

QY 2151 CACCATTTGTGAAGAAACATTAACCAAGTTAAATGATGTGCGGAGTTT----- 2204  
 DB 1133 CACCATTTGTGAAGAAACATTAACCAAGTTAAATGATGTGCTGAGTTT----- 1192

QY 2205 -----TTTCTCTCTTTTAAATGCTGCTTAAAGCTT 2241  
 DB 1193 TGTGTGTTTGTGATGATGCTGCTTTTCTTTTAAAGTGTGTTTAAAGCT 1252

QY 2242 TAAACAATGGAAACCCCTGTGAGACATGCTGATATCTTGTGAGAAACCAAGAGGCTC 2301  
 DB 1253 TAAACAATGGAAACCCCTGTGAGACATGCTGATATCTTGTGAGAAACCAAGAGGCTC 1312

QY 2302 TTAAGTGAACAAATGTCATGTTGTGAGTGTGTTTAAATTAATTTCA 2361  
 DB 1313 -TAATGTAACAAATGTCATGTTGTGAGTGTGTTTAAATTAATTTCA 1362

QY 2362 AATGTTAATAA 2373  
 DB 1363 AATGTTAATAA 1374

RESULT 15  
 BX325713 1049 bp mRNA linear EST 08-APR-2004  
 LOCUS BX325713 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens CDNA  
 DEFINITION clone CS0D1039YJ24 5-PRIME, mRNA sequence.  
 ACCESSION BX325713  
 VERSION BX325713.2 GI:46280663  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 1049)  
 AUTHORS Li, W.B., Gruber, C., Jessee, J. and Polyes, D.  
 TITLE Full-length CDNA libraries and normalization  
 JOURNAL Unpublished (2001)  
 COMMENT On May 2, 2003 this sequence version replaced gi:30344479.  
 CONTACT: Genoscope  
 Genoscope - Centre National de Sequenage  
 2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE  
 Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr  
 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime  
 end enriched, double-strand cDNA was digested with Not I and cloned  
 into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library  
 was normalized. Library was constructed by Life Technologies, a  
 division of Invitrogen. This sequence belongs to sequence cluster  
 10589.f  
 For more information about this cluster, see  
 http://www.genoscope.cns.fr/cdna?c=CS0A5012H20P1&c=10589.f.  
 Location/Qualifiers  
 1..1049  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="CS0D1039YJ24"  
 /issue\_type="PLACENTA COT 25-NORMALIZED"  
 /note="1st strand cDNA was primed with a NotI-oligo(dT)  
 primer. Five prime end enriched, double-strand cDNA  
 was digested with Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized."

Query Match 41.4%; Score 989.8; DB 5; Length 1049;



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 20, 2005, 00:41:37, Search time 8501.87 Seconds  
(without alignments)  
11541.209 Million cell updates/sec

Title: US-10-791-017A-1\_COPY\_8\_2032

Sequence: 1 gacggcagcttgagagaaga.....agatgcagagaccgcgcaga.2025

Scoring table: IDENTITY\_NUC  
Gapop 10.0, Gapext 1.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database:

GenBank1: \*  
1: gb\_ba: \*  
2: gb\_ncg: \*  
3: gb\_in: \*  
4: gb\_cm: \*  
5: gb\_ov: \*  
6: gb\_pat: \*  
7: gb\_ph: \*  
8: gb\_pl: \*  
9: gb\_pr: \*  
10: gb\_ro: \*  
11: gb\_ste: \*  
12: gb\_sy: \*  
13: gb\_un: \*  
14: gb\_vl: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2025	100.0	2390	6	CQ0867364
2	2023.4	99.9	2390	6	AX411125
3	2023.4	99.9	2390	6	HSEWS
4	2011.4	99.3	2371	6	A36460
5	2011.4	99.3	2371	6	AR080100
6	2008	99.2	2182	9	BC004817
7	2006	99.1	2189	9	CR456490
8	1979	97.7	2164	9	BC072442
9	1975.4	97.6	2364	9	BC011048
10	1969.4	97.3	1971	12	BT007796
11	1966	97.1	2189	9	AK056309
12	1964.4	97.0	2326	6	C0730352
13	1884.4	93.1	193559	2	AC137500
14	1879.6	92.8	155815	2	AC018774
15	1879.6	92.8	180718	9	AL596087
16	1879.6	92.8	182501	2	AC011221
17	1798	88.8	2189	10	BC068226
18	1787.2	88.3	2188	6	AX305537
19	1787.2	88.3	2188	10	MMEWS

20	1785.6	88.2	2396	5	BC083960
21	1731.6	85.5	132906	2	AC109802
22	1731.6	85.5	163104	2	AC110672
23	1652.4	81.6	1988	6	AX714390
24	1652.4	81.6	1988	6	AK056681
25	1625.8	80.3	2026	6	C0850483
26	1625.8	80.3	2026	6	AK127624
27	1517.6	74.9	226400	2	AC128482
28	1517.6	74.9	235241	2	AC098231
29	1517.6	74.9	263925	2	AC106522
30	1428	70.5	1783	6	C0721057
31	1383.4	67.3	221444	2	AC113313
32	1373.8	67.8	247757	2	AC121282
33	1302.4	64.3	2440	5	AJ719366
34	1223	60.4	221924	2	AC105889
35	1218	60.1	275105	2	AC118907
36	1168.8	57.7	270337	2	AC099212
37	1014.4	50.1	7293	9	HSMB08920
38	1010.8	49.9	1303	9	BC000527
39	979	48.3	1864	9	AF254086
40	979	48.3	2179	9	AF254087
41	976.4	48.3	2252	9	AF254088
42	976.4	48.2	2473	9	AK026270
43	829.4	41.0	2107	9	HSU35622
44	819	40.4	2412	6	I66247
45	801.2	39.6	2665	5	BC063928

# ALIGNMENTS

RESULT 1  
LOCUS CQ0867364  
DEFINITION Sequence 1 from Patent EP1455190.  
ACCESSION CQ0867364  
VERSION CQ0867364.1 GI:51997589  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
REFERENCE 1  
AUTHORS Odendorf, M. and Wolf, S.  
TITLE Modulation of the activity of nuclear receptors via EMS  
JOURNAL Patent: EP 1455190-A 1 08-SEP-2004;  
Schering Aktiengesellschaft (DE)  
FEATURES  
Source location/Qualifiers  
1..2390  
/organism="Homo sapiens"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:9606"  
44..2014  
/note="unnamed protein product"  
/codon\_start=1  
/protein\_id="CAH33891.1"  
/db\_xref="GI:51997590"  
GOPTDVSYTOATTATYGGTAAVATSGAOPGTGTTPTAPDAVSGAPVGGTGAADTT  
/translation="MASTDVSTYSQAHAOCYSAYTAOPTGQAOTTAAYGQOYSTGY  
GPTDVSYTOATTATYGGTAAVATSGAOPGTGTTPTAPDAVSGAPVGGTGAADTT  
ATTTTQASVTAQASATGCTPAYAGGQGAAPATPMPDQGNKTFESQPPSSGYNQ  
PSLGTQASVSTYQVPGSTPMQDPTVATPSTYPTSTYSDTSSTIDQSTSDQNTYGGPS  
STGQSSVYQOQSSVYQOQPTSPYPTQSTYSQAASVYSDSSYQOQSSFPQDHPSSNG  
VYQOQSSVYQOQSSVYQOQPTSPYPTQSTYSQAASVYSDSSYQOQSSFPQDHPSSNG  
KPGQPMDEGPDLDLGPVDPEDSDNSATLYVQGLNDSVTLIDLADFPKQGVYQANR  
TQGPQPMHLYLDEKTRKGDATVSYSDPTAKAAVWFPQKDGSLKLYLARKKPP  
NMSMRGLP.PRRGRGMPPLRGGRGPGGPGGPGMGPGMGCGDGRGPPRGRGRGN  
PBGCAVQRAEDMOCNPGCCGNQNTAMTECHQCAVYPRBGLTLPFPPPPPGDNRGRG  
GPGMRGSGGLMDRGCGPMFPGSGSGSGRGRGGRGMDRGFGGGRGPGGPGCPGP  
LMEQWGRGRGGRGPGKMDKGBRHQRRDRPY"

# ORIGIN

Query Match 100.0%; Score 2025; DB 6; Length 2390;  
Best Local Similarity 100.0%; Pred. No. 0;







Accession	Sequence	Year
Db 1668	GGCTTTGTGTGGAGGAGCGAGTGCGCCCTGTGGACCTTTGATGAAACAG	1927
Qy 1221	ATGGGAGGAAAGAAGGAGGACGCGAGGACACTGGAAAAAAGGATTAAGCGGACACCGT	1980
Db 1328	ATGGGAGGAAAGAAGGAGACGCGAGGACCTGGAAAAAAGGATTAAGCGGACACCGT	1987
Qy 1981	CAGGAGCGCAGAGATCGGCCCTACTAGATGAGAGACCCCGGACA	2025
Db 1988	CAGGAGCGCAGAGATCGGCCCTACTAGATGAGAGACCCCGGACA	2032

LOCUS	2390 bp	mRNA	linear	PRI 28-JUN-1995
HSEMS				
HSEMS				
RESULT 3				

DEFINITION	H.sapiens EMS mRNA.
ACCESSION	X66899
VERSION	X66899.1 GI:547565
KEYWORDS	RNA binding protein.
SOURCE	Homo sapiens (human)
ORGANISM	Homo sapiens

REFERENCE  
AUTHORS  
1  
Delattre, O., Zucman, J., Plougairel, B., Desmazes, C., Melot, T.,

**TITLE** Gene fusion with an ETS DNA-binding domain caused by chromosome translocation in human tumours  
**JOURNAL** *Nature* 359 (6391), 162-165 (1992)

PUBMED REFERENCE AUTHORS TITLE JOURNAL COMMENT	
1522903 2 (bases 1 to 2371) Delattre, O. Direct Submission Submitted (26-MAY-1992) O. Delattre, Lab. de Genet. des Tumeurs. Inst. Curie, 26 rue D'Ulm, 75331 Paris Cedex, FRANCE On Sep 23, 1994 this sequence version replaced gi:31279.	

COMMENT	On Sep 23, 1994 this sequence version replaced gl:31279.
FEATURES	Location/Qualifiers
SOURCE	1. .2390

```

/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="caxon:9606"
/chromosome="22q12"
/clone="Bp1AC5"
/tissue_type="brain"
/clone_id="cDNA, Stratagene 936206"
/dev_stage="fetus"
1. .2390
gene
/gene="EWS"
44. .2014
CDS

```

```

/ gene="BMS"
/ codon_start=1
/ protein_id="RNA binding protein"
/ db_xref="GI:31280"
/ db_xref="GOA:Q01844"
/ db_xref="UniProt/Swiss-Prot:Q01844"
/ translation="MASTDITVYSOAAAGQGSATVTPATGCGAATTAAGAGQGSXGTV"
GAPDVSYTAOATATVGGTAVATVSYGDPVDDSDNSALVYGLADSVLIDLADPFCOCGVKKKKR
ATVTTTQASVYAGQSAVGTOPATVPAVGGQPAATATPRPDGKPTETSGQSTGTYN
PSLGGQGSNSTYSPQVPSYPMQVPTAPSPYPTSTYSVTPSTYDSSSTQQNTQGP
SYGQGSYVGQOSSYVGQOPTSYYPQTGYSQAPSYSGQOSSYSYGQSSFRDHPBSMQ
VYGGSGSGFSGPGENRSMGSPNMRGSGFGRGMSGSGGSGGMSGGERGSGFGRG
KPGSGMDSGPDLIDGAPVDVDDSDNSALVYGLADSVLIDLADPFCOCGVKKKKR
TGCPNHTIYLDERTCKPKGDATVSYEDPPTAKAANEMDKDFOGSKTKVSLARKPP
MNSMKGGLPREGKMPPLPLRGSGGPGGSGPMKSGRGGRGDRGFRPPRGSRGK
PSGGGVNVRAGADWQCPNPGCGNQNPAMRTENOCAPKAPREGFLPPPPPGSGRG
GPGCGRGGRGGLMDRGPGGMFRGGRGDRGGRGGRGDRGGRGGRGGRGGRGGRG
LMEQWGGRRGGRGGRGPKMDKGBHROGERDRPY"
1127 1387
/ bound_molecule="RNA"
/ bound_molecule="BMS"

```

```

polyA_signal      2162.  .2167
                  /gene="EWS"
                  /evidence=experimental
polyA_signal      2350.  .2355
                  /gene="EWS"
                  /evidence=experimental
polyA_site        2371
                  /gene="EWS"
ORIGIN

```

Query Match	99.9%	Score	2023.4	DB	9	length	2390
Best Local Similarity	100.0%	Pred.	No. 0				
Matches 2024	0	Mismatches	1	Indels	0	Gaps	0

OY	1	GACGGACGTTGAGAGAA	CGAGAGAGAGAGAGAA	AAATGACGTC	TCACGGAA	TTACAG	TAAC	120	
Db	8	GACGACGTTGAGAGAA	CGAGAGAGAGAGAA	AAATGACGTC	TCACGGAA	TTACAG	TAAC	67	
OY		61	TATATGCCAAGCTGAGAGCC	CAGACAGGGCTTA	CAGTGCCTTA	CAACCCGCCAGCCCA	CTCAAGGA	120	
Db		68	TATATGCCAAGCTGAGAGCC	CAGACAGGGCTTA	CAGTGCCTTA	CAACCCGCCAGCCCA	CTCAAGGA	127	
OY		121	TATGCACAGACCAACCC	CAGGACATATGGGCA	CAAAAGCTATGGA	ACTTATGCA	CAGCCCACT	180	
Db		128	TATGCACAGACCAACCC	CAGGACATATGGGCA	CAAAAGCTATGGA	ACTTATGCA	CAGCCCACT	187	
OY		181	GATGTCAAGCTATATCC	AGAGGCTCAGACCA	CTTATGGGCA	GAGACCCGCTATGCA	ACT	240	
Db		188	GATGTCAAGCTATATCC	AGAGGCTCAGACCA	CTTATGGGCA	GAGACCCGCTATGCA	ACT	247	
OY		241	TCTTATGAGACAGCCTCC	CACTGTTTAACTCA	CTGACGCCCCCA	GGCATA	CAGACGAG	300	
Db		248	TCTTATGAGACAGCCTCC	CACTGTTTAACTCA	CTGACGCCCCCA	GGCATA	CAGACGAG	307	
OY		301	CCTGTCCAGGGGGTATG	GCATGTGTCTTATGAT	ACAACA	CTGCTTACAGT	CACCAACACC	360	
Db		308	CCTGTCCAGGGGGTATG	GCATGTGTCTTATGAT	ACAACA	CTGCTTACAGT	CACCAACACC	367	
OY		361	CAGGCCCTCTTATGAG	ACTGTCTGCA	TATGGGCA	CTCAGCCTGTCTTAT	CCAGGCTATGAGG	420	
Db		368	CAGGCCCTCTTATGAG	ACTGTCTGCA	TATGGGCA	CTCAGCCTGTCTTAT	CCAGGCTATGAGG	427	
OY		421	CAGAGGCGAGGACCACT	GCACCTTACAAGA	CCGAGAGATGGAAA	CAAGCCCACTGAGA	CT	480	
Db		428	CAGAGGCGAGGACCACT	GCACCTTACAAGA	CCGAGAGATGGAAA	CAAGCCCACTGAGA	CT	487	
OY		481	AGTCAACTCTAATCTAG	ACA	CAGGGGGTTTACA	CAACGAGCC	CAGCTTAGATATGACA	GAAGT	540
Db		488	AGTCAACTCTAATCTAG	ACA	CAGGGGGTTTACA	CAACGAGCC	CAGCTTAGATATGACA	GAAGT	547
OY		541	AACCTACAGTTATCCCA	GAGTACCTGGGAGCTAC	CCCAATGACAGGCTCA	CTGCACCTTCA		600	
Db		548	AACCTACAGTTATCCCA	GAGTACCTGGGAGCTAC	CCCAATGACAGGCTCA	CTGCACCTTCA		607	
OY		601	TCTTACCCCTCTTAC	AGCTATTCCTCTTACA	CAGGCCCA	CTAGTTATGAT	CAGAGCAGTTAC	660	
Db		608	TCTTACCCCTCTTAC	AGCTATTCCTCTTACA	CAGGCCCA	CTAGTTATGAT	CAGAGCAGTTAC	667	
OY		661	TCTCAGACGAACA	CACTTATGGGCAAC	CGAGAGCTATGGA	CAGGAGAGTATG	GTCTCA	720	
Db		668	TCTCAGACGAACA	CACTTATGGGCAAC	CGAGAGCTATGGA	CAGGAGAGTATG	GTCTCA	727	
OY		721	CAAGACAGCTATGGG	CAGACGCTTCCCA	CTAGTTACCA	CCCCCAACTGAT	TCTTCA	780	
Db		728	CAAGACAGCTATGGG	CAGACGCTTCCCA	CTAGTTACCA	CCCCCAACTGAT	TCTTCA	787	
OY		781	CAAGCTCCAACTCA	TATATGCAAC	AGAGCAGCAGCTA	CCGGCAGCAGAGTTCAT	TCCGA	840	
Db		788	CAAGCTCCAACTCA	TATATGCAAC	AGAGCAGCAGCTA	CCGGCAGCAGAGTTCAT	TCCGA	847	
OY		841	CAGAGCAACCCCA	GTAGATGAGTAT	TATGGGAGAGGCTG	GGAGATTTTCCG	AGCA	900	
Db		848	CAGAGCAACCCCA	GTAGATGAGTAT	TATGGGAGAGGCTG	GGAGATTTTCCG	AGCA	907	

QY 901 GGAGAGAACCGGAGCATAGATGGCCCTGATTAACCGGGGCAAGGGAAGAGGGGATTTGAT 960  
DB 908 GGAGAGAACCGGAGCATAGATGGCCCTGATTAACCGGGGCAAGGGAAGAGGGGATTTGAT 967  
QY 961 CGTGAAGCATAGACAGAGGTGGCGGGAGAGAGACGCGGTGATATGGCAGCGCTGA 1020  
DB 968 CGTGAAGCATAGACAGAGGTGGCGGGAGAGAGACGCGGTGATATGGCAGCGCTGA 1027  
QY 1021 GAGCGAGGTGGCTTCAATTAAGCTGTGACCCATGATGAAGACCAATCTTGTATCTA 1080  
DB 1028 GAGCGAGGTGGCTTCAATTAAGCTGTGACCCATGATGAAGACCAATCTTGTATCTA 1087  
QY 1081 GGGCCCACTGTATGATCCAGATGAAGACTCTGACCAACAGTGAATTTATGACAGATTA 1140  
DB 1088 GGGCCCTCTGTATGATCCAGATGAAGACTCTGACCAACAGTGAATTTATGACAGATTA 1147  
QY 1141 AATGACAGTGTGATCTAGATGATCTGGCAGACTTCTTTAAGCAGTGTGGGGTTGTAAAG 1200  
DB 1148 AATGACAGTGTGATCTAGATGATCTGGCAGACTTCTTTAAGCAGTGTGGGGTTGTAAAG 1207  
QY 1201 ATGAACAGAGAACTGGGCAACCCATGATCACTACTGGAACAAGAAACAGAAAG 1260  
DB 1208 ATGAACAGAGAACTGGGCAACCCATGATCACTACTGGAACAAGAAACAGAAAG 1267  
QY 1261 CCCAAAGGCGATGCCAAGTGTCTATGAAGACCCCACTGCCAAGGCTGCCGTGGA 1320  
DB 1268 CCCAAAGGCGATGCCAAGTGTCTATGAAGACCCCACTGCCAAGGCTGCCGTGGA 1327  
QY 1321 TGGTTTGAATGGGAAGATTTTCAAGGGAGCAAACTTAAAGTCCCTTGTCTGGAGAAAG 1380  
DB 1328 TGGTTTGAATGGGAAGATTTTCAAGGGAGCAAACTTAAAGTCCCTTGTCTGGAGAAAG 1387  
QY 1381 CTTCCAAATGAACGATATGCGGGGTGTCTGCAACCCCGTAGAGGACAGAGGATGCCACA 1440  
DB 1388 CTTCCAAATGAACGATATGCGGGGTGTCTGCAACCCCGTAGAGGACAGAGGATGCCACA 1447  
QY 1441 CCACTCCGTGAGAGTCCAGAGAGGCCCGAGAGGTCTGTGGGGGACCCATGGGTCCGATGGGA 1500  
DB 1448 CCACTCCGTGAGAGTCCAGAGAGGCCCGAGAGGTCTGTGGGGGACCCATGGGTCCGATGGGA 1507  
QY 1501 GGCCTGAGAGGATAGAGAGGCTTCCCTCAAGAGAGACCCCGGGGTTCGCGAGGGAGAC 1560  
DB 1508 GGCCTGAGAGGATAGAGAGGCTTCCCTCAAGAGAGACCCCGGGGTTCGCGAGGGAGAC 1567  
QY 1561 CCTCTGAGAGGAGAAAGTCCAGACCGAGCTGAGAGCTGCAATGTCTCCAAATCCGGGT 1620  
DB 1568 CCTCTGAGAGGAGAAAGTCCAGACCGAGCTGAGAGCTGCAATGTCTCCAAATCCGGGT 1627  
QY 1621 TGTGGAACCAAGAACTTCCCTGAGAAACAGAGTGCACCAAGTGTAAAGGCCCGCAAGGCT 1680  
DB 1628 TGTGGAACCAAGAACTTCCCTGAGAAACAGAGTGCACCAAGTGTAAAGGCCCGCAAGGCT 1687  
QY 1681 GAAGGCTTCTCCCGCAACCTTTCGCGCCCGGGGTGTATGTGGCAGAGGTGGCCCT 1740  
DB 1688 GAAGGCTTCTCCCGCAACCTTTCGCGCCCGGGGTGTATGTGGCAGAGGTGGCCCT 1747  
QY 1741 GGTGAGCATGCGGAGAGAAAGGTGGCTCATGATGATGTGTGGTCCCGGTGAAATGTTTC 1800  
DB 1748 GGTGAGCATGCGGAGAGAAAGGTGGCTCATGATGATGTGTGGTCCCGGTGAAATGTTTC 1807  
QY 1801 AAGAGTGGCCGTGTGTGAGACAGAGTGTCTCCGTGTGGTGGCGGGACATGAACCGAGGT 1860  
DB 1808 AAGAGTGGCCGTGTGTGAGACAGAGTGTCTCCGTGTGGTGGCGGGACATGAACCGAGGT 1867  
QY 1861 GGCCTTGTGTGAGAGAAAGAGGTGGCTTTCGCGGGGGGCCCTGTGAACCTTTATGAGAAACAG 1920  
DB 1868 GGCCTTGTGTGAGAGAAAGAGGTGGCTTTCGCGGGGGGCCCTGTGAACCTTTATGAGAAACAG 1927  
QY 1921 ATGGAGAGAAAGAGAGAGAGCTGTGAGACCTGTGAAATTAAGTAAAGCGAGCACCGGT 1980  
DB 1928 ATGGAGAGAAAGAGAGAGAGCTGTGAGACCTGTGAAATTAAGTAAAGCGAGCACCGGT 1987

QY 1981 CAGAGGCGCAGAGATGGCCCTACTAGATCAGAGACCCCGCAGA 2025  
DB 1988 CAGAGGCGCAGAGATGGCCCTACTAGATCAGAGACCCCGCAGA 2032

RESULT 4  
A36460  
LOCUS A36460 2371 bp DNA linear PAT 05-MAR-1997  
DEFINITION Sequence 1 from Patent WO9323549.  
ACCESSION A36460  
VERSION A36460.1 GI:2293778  
KEYWORDS  
SOURCE  
ORGANISM  
Homo sapiens (human)  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE  
Aurias A., Delattre O., Desmazes C., Melot T., Peter M.,  
1 (bases 1 to 2371)  
NUCLEIC ACID CORRESPONDING TO A GENE OF CHROMOSOME 22 INVOLVED IN  
RECURRENT CHROMOSOMAL TRANSLOCATIONS ASSOCIATED WITH THE  
DEVELOPMENT OF CANCEROUS TUMORS  
Patent: WO 9323549-A 1 25-NOV-1993;  
CENTRE NAT RECH SCIEN (FR)  
Other publication FR 2691475 931126  
Other publication JP 85009647 960206.  
COMMENT  
Other publication FR 2691475 931126  
Other publication JP 85009647 960206.  
FEATURES  
source  
1. 2371  
Location/Qualifiers  
/organism="Homo sapiens"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:9606"  
/dev\_stage="FOETUS"

ORIGIN  
Query Match 99.3%; Score 2011.4; DB 6; Length 2371;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 2012; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 13 GAGAACGAGAGAGAGAGAAATGGCGTCCAGGATTTACAGTACTATAGCAAGCT 72  
DB 1 GAGAACGAGAGAGAGAGAAATGGCGTCCAGGATTTACAGTACTATAGCAAGCT 60  
QY 73 GCAGCCGACAGAGGCTACAGTGTCTTACACCGCCCAAGCTCAAGATATGACAGACC 132  
DB 61 GCAGCCGACAGAGGCTACAGTGTCTTACACCGCCCAAGCTCAAGATATGACAGACC 120  
QY 133 ACCCAGGCAATATGGGCAACAAAGCTATGGAACCTATGACAGGCCACTGATGCAGTAT 192  
DB 121 ACCCAGGCAATATGGGCAACAAAGCTATGGAACCTATGACAGGCCACTGATGCAGTAT 180  
QY 193 ACCCAGGCTCAGACCACTGCAACTATGGGACAGCCGCTATGCAACTTCTTATGACAG 252  
DB 181 ACCCAGGCTCAGACCACTGCAACTATGGGACAGCCGCTATGCAACTTCTTATGACAG 240  
QY 253 CTTCCCACTGTATATCTACTCACTGCACTGCCCCCAGGACATACAGCAGCTGTCCAGGG 312  
DB 241 CTTCCCACTGTATATCTACTCACTGCACTGCCCCCAGGACATACAGCAGCTGTCCAGGG 300  
QY 313 TATGACACTGTGTCTATATGATACCACTGCTACTGATGACAGCCACCAAGGCTCTCTAT 372  
DB 301 TATGACACTGTGTCTATATGATACCACTGCTACTGATGACAGCCACCAAGGCTCTCTAT 360  
QY 373 GCAGCTCAGTCTGATATGAGCACTCAGCTGCTTATCAGACCTATGGGAGCAGCCAGCA 432  
DB 361 GCAGCTCAGTCTGATATGAGCACTCAGCTGCTTATCAGACCTATGGGAGCAGCCAGCA 420  
QY 433 GCGACTGCACTTCAAGAGCCGAGAGTGAACCAAGCCCACTGAGACTGTCAACTCA 492  
DB 421 GCGACTGCACTTCAAGAGCCGAGAGTGAACCAAGCCCACTGAGACTGTCAACTCA 480  
QY 493 TCTAGCAGAGGGGGTTACCAACGAGCCAGCTAGATATGACAGAGTAACTACAGTAT 552  
DB 481 TCTAGCAGAGGGGGTTACCAACGAGCCAGCTAGATATGACAGAGTAACTACAGTAT 540

OY	553	CCCCAGGACTCTGGAGAGCTACCCCATGAGCAGCACTGCACCTCCATCCTACCTCT	612
Db	541	CCCCAGGACTCTGGAGAGCTACCCCATGAGCAGCACTGCACCTCCATCCTACCTCT	600
OY	613	ACCAAGCTATTCCTCTACACAGCCGACCTAGTTATGATCAAGACAGTACTCTCAGCAGAAC	672
Db	601	ACCAAGCTATTCCTCTACACACCCCACTAGTTATGATCAAGACAGTACTCTCAGCAGAAC	660
OY	673	ACCAATGGGCAACCGAGAGCTATGAGCAGAGAGTACGTATGGTCAACAAAGACGTAT	732
Db	661	ACCAATGGGCAACCGAGAGCTATGAGCAGAGAGTACGTATGGTCAACAAAGACGTAT	720
OY	733	GGGAGCAGCCTCCCACTAGTTAACCCCAACTGGATCCTACAGCCAGTCTCAAGT	792
Db	721	GGGAGCAGCCTCCCACTAGTTAACCCCAACTGGATCCTACAGCCAGTCTCAAGT	780
OY	793	CAATATGACCAACAGACGACGACTACGGGACGACAGTTCATTTCGACAGAACCAACCC	852
Db	781	CAATATGACCAACAGACGACGACTACGGGACGACAGTTCATTTCGACAGAACCAACCC	840
OY	853	AGTAGCAGTGGGTGTTATGGGCAAGAGCTGAGAGATTTTTCGGACAGAGAAACCGG	912
Db	841	AGTAGCAGTGGGTGTTATGGGCAAGAGCTGAGAGATTTTTCGGACAGAGAAACCGG	900
OY	913	AGCATGAGTGGCCCTGATTAACCGGGGACAGGGGAAAGGGGGATTTGATCGTAGAGCATG	972
Db	901	AGCATGAGTGGCCCTGATTAACCGGGGACAGGGGAAAGGGGGATTTGATCGTAGAGCATG	960
OY	973	AGCAGAGGTGGGGCGGGGAGAGACCGCGGTGAATGGGCAACGCGTCGAGACGAGGTGGC	1032
Db	961	AGCAGAGGTGGGGCGGGGAGAGACCGCGGTGAATGGGCAACGCGTCGAGACGAGGTGGC	1020
OY	1033	TTCAATAAGCCTGTGTGAACCCATGGATGAAGAACCAATCTTGATCTTAGGCCCACTGTA	1092
Db	1021	TTCAATAAGCCTGTGTGTGAACCCATGGATGAAGAACCAATCTTGATCTTAGGCCCACTGTA	1080
OY	1093	GATCCAGATGAAGACTCTGACCAACAGTGCATTTATGTACAAAGATTTAAATGACAGTGTG	1152
Db	1081	GATCCAGATGAAGACTCTGACCAACAGTGCATTTATGTACAAAGATTTAAATGACAGTGTG	1140
OY	1153	ACTCTAGATGATCTGGCAGACTCTTTTAAGCAGTGTGGGGTTGTTAAGATGAACAAGAGA	1212
Db	1141	ACTCTAGATGATCTGGCAGACTCTTTTAAGCAGTGTGGGGTTGTTAAGATGAACAAGAGA	1200
OY	1213	ACTGGGCAACCCATGATCAACTTACCTTGACCAAGAAACAGAAACCCCAAAAGCGAT	1272
Db	1201	ACTGGGCAACCCATGATCAACTTACCTTGACCAAGAAACAGAAACCCCAAAAGCGAT	1260
OY	1273	GCACAGTGTCTTATGAAGACCCACCACTGCACAGAGCTGCGTGGAAATGGTTGATGGG	1332
Db	1261	GCACAGTGTCTTATGAAGACCCACCACTGCACAGAGCTGCGTGGAAATGGTTGATGGG	1320
OY	1333	AAAGATTTTCAAGGGAGCAAACTTAAAGTCTCCCTTGTCTGGAAAGAACCTTCAATGAAC	1392
Db	1321	AAAGATTTTCAAGGGAGCAAACTTAAAGTCTCCCTTGTCTGGAAAGAACCTTCAATGAAC	1380
OY	1393	AGTATATCGGGGGTGTCTGCGCAACCCCTGTAGGGCAGAGGCAATGCCACCACTCCGTGGA	1452
Db	1381	AGTATATCGGGGGTGTCTGCGCAACCCCTGTAGGGCAGAGGCAATGCCACCACTCCGTGGA	1440
OY	1453	GGTCCAGAGAGGCCAGAGAGTCTTGGGGGAAACCATGGGTGCATGGAGGCGCGTAGAGGA	1512
Db	1441	GGTCCAGAGAGGCCAGAGAGTCTTGGGGGAAACCATGGGTGCATGGAGGCGCGTAGAGGA	1500
OY	1513	GATAGAGAGGCTTCTCTTCAAGAGAGACCCCGGGGTTCGCCAGGGAACCCCTCTGAGAGA	1572
Db	1501	GATAGAGAGGCTTCTCTTCAAGAGAGACCCCGGGGTTCGCCAGGGAACCCCTCTGAGAGA	1560
OY	1573	GGAAACGTCCAGACCCAGAGCTGGGAACCTGGCAGTGTCCCAATCCGGGTGTGGAAACAG	1632
Db	1561	GGAAACGTCCAGACCCAGAGCTGGGAACCTGGCAGTGTCCCAATCCGGGTGTGGAAACAG	1620

QY	1633	AACTTCCCTCGAGAAACAGAGTGCACCAAGTGTAAAGGCCCTGAAGCTTCCTC	1692
Db	1621	AACCTTCCCTCGAGAAACAGAGTGCACCAAGTGTAAAGGCCCTGAAGCTTCCTC	1680
QY	1693	CCGCGACCTTTCCGCCCGCCGGGTGATCGTGGCAGAGTGGCCCTTGTCATCCGG	1752
Db	1681	CCGCGACCTTTCCGCCCGCCGGGTGATCGTGGCAGAGTGGCCCTTGTCATCCGG	1740
QY	1753	GGAGGAAGAGTGGCTCATAGATCGTGGTGGTCCCGGTGGAAATGTTCAAGAGTGGCGT	1812
Db	1741	GGAGGAAGAGTGGCTCATAGATCGTGGTGGTCCCGGTGGAAATGTTCAAGAGTGGCGT	1800
QY	1813	GGTGAGACAGAGGTGGCTTCCGTTGGTGGCCCGGGGCATGGAACCGAGGTGGCTTTGGTGA	1872
Db	1801	GGTGAGACAGAGGTGGCTTCCGTTGGTGGCCCGGGGCATGGAACCGAGGTGGCTTTGGTGA	1860
QY	1873	GGAAGACGAGGTGGCCCTGGGGGGGCCCTTGAGACTTTGATGGAACGATGGAGGAAGA	1932
Db	1861	GGAAGACGAGGTGGCCCTGGGGGGGCCCTTGAGACTTTGATGGAACGATGGAGGAAGA	1920
QY	1933	AGAGAGAGAAGTGGAGGACCTTGGAATAATGTATTAAGCGACGACCGTCAAGAGCGGAGA	1992
Db	1921	AGAGAGAGAAGTGGAGGACCTTGGAATAATGTATTAAGCGACGACCGTCAAGAGCGGAGA	1980
QY	1993	GATCGGCCCTTACTGATGCAAGACCCCGCAGA 2025	
Db	1981	GATCGGCCCTTACTGATGCAAGACCCCGCAGA 2013	

```

RESULT 5
AR080100 LOCUS AR080100 2371 bp DNA linear PAT 31-AUG-2000
DEFINITION Sequence 1 from patent US 5968734.
ACCESSION AR080100
VERSION AR080100.1 GI:10006835
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS Autilas,A., Delattre,O., Desmazes,C., Melot,T., Peter,M.,
Ploogastel,B., Thomas,G. and Zucman,J.
TITLE Nucleic acid corresponding to a gene of chromosome 22 involved in
recurrent chromosomal translocations associated with the
development of cancerous tumors, and nucleic acids of fusion
resulting from said translocations
Patent: US 5968734-A 1 19-OCT-1999;
JOURNAL Location/Qualifiers
FEATURES
source 1..2371
/organism="unknown"
/mol_type="unassigned DNA"

```

	Query Match	Similarity	99.3%; Score 2011.4;	DB 6;	Length 2371;
	Best Local	Similarity	100.0%;	Pred. No. 0;	
	Matches 2012;	Conservative	0;	Mismatches 1;	Indels 0; Gaps 0;
QY	13	GAGAACGAGGAGGAAGAAGAAAATGGCGCTCCACGATTACAGTACTTATGCCAAGCT	72		
DB	1	GAGAAACGAGGAGGAAGAAGAAAATGGCGCTCCACGATTACAGTACTTATGCCAAGCT	60		
QY	73	GCAGCGCAGCAGGGCTACGTGCTTACAACCGCCCAAGCCCACTCAAGAGTATGACACAGCC	132		
DB	61	GCAGCGCAGCAGGGCTACGTGCTTACAACCGCCCAAGCCCACTCAAGAGTATGACACAGCC	120		
QY	133	ACCACGAGCATATGGGCAACAAGCTATGGAACCTATGGACAGCCCACTGATGTCAAGCTAT	192		
DB	121	ACCACGAGCATATGGGCAACAAGCTATGGAACCTATGGACAGCCCACTGATGTCAAGCTAT	180		
QY	133	ACCACGAGCTACAGCACATGCAACCTATGGGGAGACCGCTATGGCAACTTTTATGGAAG	252		
DB	181	ACCACGAGCTACAGCACATGCAACCTATGGGGAGACCGCTATGGCAACTTTTATGGAAG	240		

QY 253 CCTCCACGTGGTATATCTCACTCACTGCCCGGAGATACAGCAGCGCTGTCCAGAGG 312  
Db 241 CTTCCCACTGGTATATCTCACTCACTGCCCGGAGATACAGCAGCGCTGTCCAGAGG 300  
QY 313 TATGGCACTGGTGTATATGATACCACTGCTCACTGACCAACACCGAGCTCTCTAT 372  
Db 301 TATGGCACTGGTGTATATGATACCACTGCTCACTGACCAACACCGAGCTCTCTAT 360  
QY 373 GAGCTCACTGTGATATGATACCTCACTGCTCTATCTCACTGAGTGAAGAGAGAGCA 432  
Db 361 GAGCTCACTGTGATATGATACCTCACTGCTCTATCTCACTGAGTGAAGAGAGAGCA 420  
QY 433 GCCACTGCACTCACTCACTCACTGAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 492  
Db 421 GCCACTGCACTCACTCACTCACTGAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 480  
QY 493 TCTAGCAGAGGGGGTTCACACAGCCGAGCTAGAGTATGAGACAGAGTAACTACAGTTAT 552  
Db 481 TCTAGCAGAGGGGGTTCACACAGCCGAGCTAGAGTATGAGACAGAGTAACTACAGTTAT 540  
QY 553 CCCCAGGTACTGGAGAGTACCCCACTGAGCAGTCACTGCACTCTCACTCTCTCT 612  
Db 541 CCCCAGGTACTGGAGAGTACCCCACTGAGCAGTCACTGCACTCTCACTCTCTCTCT 600  
QY 613 ACCAGCTATCTCTCTCACTCACTGAGCAGTCACTGAGTGAAGAGAGAGAGAGAGAG 672  
Db 601 ACCAGCTATCTCTCTCACTCACTGAGCAGTCACTGAGTGAAGAGAGAGAGAGAGAG 660  
QY 673 ACCTATGGGCAACCGAGCAGTCACTGAGCAGAGTATGAGTCACTCACTCACTCACT 732  
Db 661 ACCTATGGGCAACCGAGCAGTCACTGAGCAGAGTATGAGTCACTCACTCACTCACT 720  
QY 733 GGGCAGCAGCTCTCCCACTGATACCCACCCCACTGATCTCACTCACTCACTCACT 792  
Db 721 GGGCAGCAGCTCTCCCACTGATACCCACCCCACTGATCTCACTCACTCACTCACT 780  
QY 793 CAATATAGCCAAAGAGCAGCAGTCACTGAGCAGAGTCACTGAGCAGAGAGAGAGAGAG 852  
Db 781 CAATATAGCCAAAGAGCAGCAGTCACTGAGCAGAGTCACTGAGCAGAGAGAGAGAGAG 840  
QY 853 AGTAGCAGTGGTGTATATGAGCAGAGTCTGAGAGATTTCCGAGCAGCAGAGAGAGAG 912  
Db 841 AGTAGCAGTGGTGTATATGAGCAGAGTCTGAGAGATTTCCGAGCAGCAGAGAGAGAG 900  
QY 913 AGCATAGTGGCCCTGATTAACCGGGGCAAGGGGAGATTTGATCTGAGAGAGAG 972  
Db 901 AGCATAGTGGCCCTGATTAACCGGGGCAAGGGGAGATTTGATCTGAGAGAGAG 960  
QY 973 AGCAGAGTGGGGGAG 1032  
Db 961 AGCAGAGTGGGGGAG 1020  
QY 1033 TTCAATAGAGCTGTGAGACCATGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1092  
Db 1021 TTCAATAGAGCTGTGAGACCATGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1080  
QY 1093 GATCCAGATGAG 1152  
Db 1081 GATCCAGATGAG 1140  
QY 1153 ACTCTGATGATGTGGCAGCTCTTAAAGCAGTGTGGGTTTAAAGATGAGAGAGAGAGAG 1212  
Db 1141 ACTCTGATGATGTGGCAGCTCTTAAAGCAGTGTGGGTTTAAAGATGAGAGAGAGAGAG 1200  
QY 1213 ACTGGGCAACCCATGATCTCACTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1272  
Db 1201 ACTGGGCAACCCATGATCTCACTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1260  
QY 1273 GCCACAGTGTCTTATGAG 1332  
Db 1261 GCCACAGTGTCTTATGAG 1320  
QY 1333 AAAGATTTTCAAGGAG 1392

Db 1321 AAAGATTTTCAAGGAG 1380  
QY 1393 AGTAGGAG 1452  
Db 1381 AGTAGGAG 1440  
QY 1453 GGTTCAG 1512  
Db 1441 GGTTCAG 1500  
QY 1513 GATAG 1572  
Db 1501 GATAG 1560  
QY 1573 GAAAG 1632  
Db 1561 GAAAG 1620  
QY 1633 AACTTGGCTGTGAG 1692  
Db 1621 AACTTGGCTGTGAG 1680  
QY 1693 CCGCAG 1752  
Db 1681 CCGCAG 1740  
QY 1753 GAG 1812  
Db 1741 GAG 1800  
QY 1813 GGTTCAG 1872  
Db 1801 GGTTCAG 1860  
QY 1873 GGAAG 1932  
Db 1861 GGAAG 1920  
QY 1933 AG 1992  
Db 1921 AG 1980  
QY 1993 GATGGCCCTTCACTGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2025  
Db 1981 GATGGCCCTTCACTGATGAG 2013

RESULT 6  
BC004817  
LOCUS  
DEFINITION  
Homo sapiens Ewing sarcoma breakpoint region 1, transcript variant  
BC004817  
ACCESSION  
VERSION  
KEYWORDS  
KEYWORDS  
ORGANISM  
Homo sapiens  
Homo sapiens  
REFERENCE  
AUTHORS  
1 (bases 1 to 2182)  
Straussberg, R.L., Feligold, E.A., Grouse, L.H., Derge, J.G.,  
Klausner, R.D., Collins, F.S., Wagner, K.H., Schaefer, C.F., Bhat, N.K.,  
Altschul, S.F., Zeeberg, B., Burow, K.H., Schaefer, C.F., Bhat, N.K.,  
Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F.,  
Ditschenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,  
Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L.,  
Schetz, T.B., Brownstein, M.J., Usdin, T.B., Toshiyuki, S.,  
Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J.,  
Abramson, R.D., Mullany, S.J., Bosak, S.A., McEwan, P.J.,  
McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S.,  
Morley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W.,  
Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,



JOURNAL  
PUBMED  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL

TITLE

12477932  
2 (pages 1 to 2182)  
Strausberg, R.  
Direct Submission  
Submitted (21-MAR-2001) National Institutes of Health, Mammalian  
Gene Collection (MGC), Cancer Genomics Office, National Cancer  
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
USA

NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
Contact: MGC help desk  
Email: [cgabs-remail.nih.gov](mailto:cgabs-remail.nih.gov)  
Tissue Procurement: ATCC  
cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Baylor College of Medicine Human Genome  
Sequencing Center  
Center code: BCM-HGSC  
Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>  
Contact: [amg@bcm.tmc.edu](mailto:amg@bcm.tmc.edu)  
Guarnante, P.H., Garcia, A.M., Lu, X., Huiyk, S.W., Louisged, H.,  
Kowle, C.R., Sneed, A.J., Martin, R.G., Muzny, D.M., Nanaavati,  
A.N., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
 Series: IRAX Plate: 3 Row: 1 Column: 6  
 This clone was selected for full length sequencing because it passed the following selection criteria: Hexamer frequency ORF analysis.

FEATURES	Location/Qualifiers
SOURCE	1. .2182

```

/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="MGC:5432 IMAGE:344915"
/tissue_type="Placenta, Chorlocarcinoma"
/clone_1db="NIH_MGC_10"
/lab_host="DH10B"
/note="Vector: pCMV-SPORT6"
1.2182
/gene="EWSR1"
/note="synonym: EWS"
/db_xref="LOCUSID:2130"
/db_xref="MIM:133450"
20.11990

```

```

/gene="BMSR1"
/codon_start=1
/product="Ewing sarcoma breakpoint region 1, isoform EMS"
/protein_id="AAH04817.1"
/db_xref="GI:13435963"
/db_xref="LOCUSID:2130"
/db_xref="MIM:133450"
/translacion="MASTPDEVSTGSAQAQCYSAVTAQPTGQAQTTOAGQSGYGTI
GQPDVSTYTAQTAQTATVGCQTAVTSYGGPFGYTPPTAPAKASQVDSQVAGADYTT
ATVTTTASVAVSQSVGTPQAPVAPVAGQAPADAPRQDPQNKPTSSQOSTGTYGN
PSLTVGQSNVSYTPVQVSGYPMQPTPLPSPSYPTVSYSSQPSYQSSSQDQNTGQPP
SYGQSSITGQOQSTGQQDPPTSYPTQTSYQAPQSTSYQOQSSYQOQSSFTQDHPBSMQ
VYQGESQGSFSGPGENRSMSPGPNRGRGSGDPRGMSQSGRGGGQSGAEGRFQSG
KPGPQSGMDGPDLDLGPVPVDPEDSDNSN1LYVQGLDNLTLTLDLDFQFCQVYKNNK
TGQSMIH1LYDKETGPKKDAIVSYEDPPTAKAAVEMWGDGDFQSKLTVLARKKQ
MNSRNGELPREGRGMPPLRGGPCGPGCGPQPMRWMDRGDGGFPPRQSRGSRG
PSSGQNYDHRGAQDQCNPPGCGQGNQNFARTCTNCKAKPKEGFLDPPPPPPGCGRGR
GPGCMGRGQGLMDRQPGCMFRGQGRGDRGCGRGMNDRGFGGGRGCGPFGC
LMEQNGRGRGGRGGKMDKGBHRRERDRLP"

```

## ORIGIN

	Query Match	99.2%; Score 2008;	DB 9;	Length 2182;	
	Best Local Similarity	100.0%;	Pred. No. 0;		
	Matches 2008;	Conservative	0;	Mismatches	0;
				Indels	0;
				Gaps	0;
QY	18	CGAGAGGAGAGAGAGAAAAATGGCGCTTCACGGATTACAGTACTTATAGCCAACTGACG	77		
Db	1	CGAGAGGAGAGAGAGAAAAATGGCGCTTCACGGATTACAGTACTTATAGCCAACTGACG	60		
QY	78	GCAGCAGGGCTTACAGTGCTTACACCGGCCGAGCCCACTCAAGATATGCAAGACCA	137		
Db	61	GCAGCAGGGCTTACAGTGCTTACACCGGCCGAGCCCACTCAAGATATGCAAGACCA	120		
QY	138	GGCATATGGGCAACAAAGCTATGGAACCTATGGAACAGCCCACTGATGCTATACCA	197		
Db	121	GGCATATGGGCAACAAAGCTATGGAACCTATGGAACAGCCCACTGATGCTATACCA	180		
QY	198	GGCTCAGAACCACTGCAACCTATGGGAGACCGGCTATGGAACCTTATGGAAGAGCTCC	257		
Db	181	GGCTCAGAACCACTGCAACCTATGGGAGACCGGCTATGGAACCTTATGGAAGAGCTCC	240		
QY	258	CACGTGTTATACTACTCACTGCGCCCGGAGATACAGCAGCCTGTCCAGGGGATAG	317		
Db	241	CACGTGTTATACTACTCACTGCGCCCGGAGATACAGCAGCCTGTCCAGGGGATAG	300		
QY	318	CACGTGCTTATGATATCCACACTGCTTACAGTCAACAACCGAGCTTCTTATGAGC	377		
Db	301	CACGTGCTTATGATATCCACACTGCTTACAGTCAACAACCGAGCTTCTTATGAGC	360		
QY	378	TCAGTCTGATATGGGCACTGAGCTGCTTATCCAGCCTATGGGAGAGAGCAGAGCAC	437		
Db	361	TCAGTCTGATATGGGCACTGAGCTGCTTATCCAGCCTATGGGAGAGAGCAGAGCAC	420		
QY	438	TGCACTTACAAAGCCGAGATGGAACAAGCCCACTGAGACTATGTCACCTCAATTAG	497		
Db	421	TGCACTTACAAAGCCGAGATGGAACAAGCCCACTGAGACTATGTCACCTCAATTAG	480		
QY	498	CACAGGGGGTTTACAACAGCCCGCCTTAGATATGGAACAGATTACTACAGTTATCCCA	557		
Db	481	CACAGGGGGTTTACAACAGCCCGCCTTAGATATGGAACAGATTACTACAGTTATCCCA	540		
QY	558	GGTACTTGGGAGCTAACCCCATGACAGCACTGCACTGCATCTCACTCTTCTTACAG	617		
Db	541	GGTACTTGGGAGCTAACCCCATGACAGCACTGCACTGCATCTCACTCTTCTTACAG	600		
QY	618	CTATTTCCTTACACAGCCGACTAGTTATGATCAGAGCACTTACTCTCAGCAAAACCTTA	677		
Db	601	CTATTTCCTTACACAGCCGACTAGTTATGATCAGAGCACTTACTCTCAGCAAAACCTTA	660		
QY	678	TGGGCAACCGAGCAGCTATGGAACAGAGAGTAGTATGCTTACAAAGACAGCTATGGGCA	737		
Db	661	TGGGCAACCGAGCAGCTATGGAACAGAGAGTAGTATGCTTACAAAGACAGCTATGGGCA	720		
QY	738	GCAGCCTCCCACTAGTTAACCCACCCCAACTGATCTTACAGCCAGCTCCAATCAATA	797		
Db	721	GCAGCCTCCCACTAGTTAACCCACCCCAACTGATCTTACAGCCAGCTCCAATCAATA	780		
QY	798	TAGCCACAGAGCAGCAGCTTACGGGAGAGAGAGTTCAATCCGAGAGACCAACCCAGTAG	857		
Db	781	TAGCCACAGAGCAGCAGCTTACGGGAGAGAGAGTTCAATCCGAGAGACCAACCCAGTAG	840		
QY	858	CATGGGATTTATATGGGAGAGAGTCTGGAGATTTTCCGAGACAGAGAGAACCCGAGCAT	917		
Db	841	CATGGGATTTATATGGGAGAGAGTCTGGAGATTTTCCGAGACAGAGAGAACCCGAGCAT	900		
QY	918	GAGTGGCCCTGATTAACCGGGGAGAGAGAGGAGATTTGATCTGTGAGGCAATGAGAG	977		
Db	901	GAGTGGCCCTGATTAACCGGGGAGAGAGAGGAGATTTGATCTGTGAGGCAATGAGAG	960		
QY	978	AGGTGGGGGGGAGAGAGAGCGGTGGAGATGGGAGAGGCTGGAGAGCAAGGTGGCTTCAA	1037		
Db	961	AGGTGGGGGGGAGAGAGAGCGGTGGAGATGGGAGAGGCTGGAGAGCAAGGTGGCTTCAA	1020		



	Chr	Start (kb)	End (kb)	Gene	Transcript	Length (nt)	GC (%)	GC3 (%)	GC3+ (%)	GC4 (%)	GC4+ (%)	GC5 (%)	GC5+ (%)	GC6 (%)	GC6+ (%)	GC7 (%)	GC7+ (%)	GC8 (%)	GC8+ (%)	GC9 (%)	GC9+ (%)	GC10 (%)	GC10+ (%)	GC11 (%)	GC11+ (%)	GC12 (%)	GC12+ (%)	GC13 (%)	GC13+ (%)	GC14 (%)	GC14+ (%)	GC15 (%)	GC15+ (%)	GC16 (%)	GC16+ (%)	GC17 (%)	GC17+ (%)	GC18 (%)	GC18+ (%)	GC19 (%)	GC19+ (%)	GC20 (%)	GC20+ (%)	GC21 (%)	GC21+ (%)	GC22 (%)	GC22+ (%)	GC23 (%)	GC23+ (%)	GC24 (%)	GC24+ (%)	GC25 (%)	GC25+ (%)	GC26 (%)	GC26+ (%)	GC27 (%)	GC27+ (%)	GC28 (%)	GC28+ (%)	GC29 (%)	GC29+ (%)	GC30 (%)	GC30+ (%)	GC31 (%)	GC31+ (%)	GC32 (%)	GC32+ (%)	GC33 (%)	GC33+ (%)	GC34 (%)	GC34+ (%)	GC35 (%)	GC35+ (%)	GC36 (%)	GC36+ (%)	GC37 (%)	GC37+ (%)	GC38 (%)	GC38+ (%)	GC39 (%)	GC39+ (%)	GC40 (%)	GC40+ (%)	GC41 (%)	GC41+ (%)	GC42 (%)	GC42+ (%)	GC43 (%)	GC43+ (%)	GC44 (%)	GC44+ (%)	GC45 (%)	GC45+ (%)	GC46 (%)	GC46+ (%)	GC47 (%)	GC47+ (%)	GC48 (%)	GC48+ (%)	GC49 (%)	GC49+ (%)	GC50 (%)	GC50+ (%)	GC51 (%)	GC51+ (%)	GC52 (%)	GC52+ (%)	GC53 (%)	GC53+ (%)	GC54 (%)	GC54+ (%)	GC55 (%)	GC55+ (%)	GC56 (%)	GC56+ (%)	GC57 (%)	GC57+ (%)	GC58 (%)	GC58+ (%)	GC59 (%)	GC59+ (%)	GC60 (%)	GC60+ (%)	GC61 (%)	GC61+ (%)	GC62 (%)	GC62+ (%)	GC63 (%)	GC63+ (%)	GC64 (%)	GC64+ (%)	GC65 (%)	GC65+ (%)	GC66 (%)	GC66+ (%)	GC67 (%)	GC67+ (%)	GC68 (%)	GC68+ (%)	GC69 (%)	GC69+ (%)	GC70 (%)	GC70+ (%)	GC71 (%)	GC71+ (%)	GC72 (%)	GC72+ (%)	GC73 (%)	GC73+ (%)	GC74 (%)	GC74+ (%)	GC75 (%)	GC75+ (%)	GC76 (%)	GC76+ (%)	GC77 (%)	GC77+ (%)	GC78 (%)	GC78+ (%)	GC79 (%)	GC79+ (%)	GC80 (%)	GC80+ (%)	GC81 (%)	GC81+ (%)	GC82 (%)	GC82+ (%)	GC83 (%)	GC83+ (%)	GC84 (%)	GC84+ (%)	GC85 (%)	GC85+ (%)	GC86 (%)	GC86+ (%)	GC87 (%)	GC87+ (%)	GC88 (%)	GC88+ (%)	GC89 (%)	GC89+ (%)	GC90 (%)	GC90+ (%)	GC91 (%)	GC91+ (%)	GC92 (%)	GC92+ (%)	GC93 (%)	GC93+ (%)	GC94 (%)	GC94+ (%)	GC95 (%)	GC95+ (%)	GC96 (%)	GC96+ (%)	GC97 (%)	GC97+ (%)	GC98 (%)	GC98+ (%)	GC99 (%)	GC99+ (%)	GC100 (%)	GC100+ (%)	GC101 (%)	GC101+ (%)	GC102 (%)	GC102+ (%)	GC103 (%)	GC103+ (%)	GC104 (%)	GC104+ (%)	GC105 (%)	GC105+ (%)	GC106 (%)	GC106+ (%)	GC107 (%)	GC107+ (%)	GC108 (%)	GC108+ (%)	GC109 (%)	GC109+ (%)	GC110 (%)	GC110+ (%)	GC111 (%)	GC111+ (%)	GC112 (%)	GC112+ (%)	GC113 (%)	GC113+ (%)	GC114 (%)	GC114+ (%)	GC115 (%)	GC115+ (%)	GC116 (%)	GC116+ (%)	GC117 (%)	GC117+ (%)	GC118 (%)	GC118+ (%)	GC119 (%)	GC119+ (%)	GC120 (%)	GC120+ (%)	GC121 (%)	GC121+ (%)	GC122 (%)	GC122+ (%)	GC123 (%)	GC123+ (%)	GC124 (%)	GC124+ (%)	GC125 (%)	GC125+ (%)	GC126 (%)	GC126+ (%)	GC127 (%)	GC127+ (%)	GC128 (%)	GC128+ (%)	GC129 (%)	GC129+ (%)	GC130 (%)	GC130+ (%)	GC131 (%)	GC131+ (%)	GC132 (%)	GC132+ (%)	GC133 (%)	GC133+ (%)	GC134 (%)	GC134+ (%)	GC135 (%)	GC135+ (%)	GC136 (%)	GC136+ (%)	GC137 (%)	GC137+ (%)	GC138 (%)	GC138+ (%)	GC139 (%)	GC139+ (%)	GC140 (%)	GC140+ (%)	GC141 (%)	GC141+ (%)	GC142 (%)	GC142+ (%)	GC143 (%)	GC143+ (%)	GC144 (%)	GC144+ (%)	GC145 (%)	GC145+ (%)	GC146 (%)	GC146+ (%)	GC147 (%)	GC147+ (%)	GC148 (%)	GC148+ (%)	GC149 (%)	GC149+ (%)	GC150 (%)	GC150+ (%)	GC151 (%)	GC151+ (%)	GC152 (%)	GC152+ (%)	GC153 (%)	GC153+ (%)	GC154 (%)	GC154+ (%)	GC155 (%)	GC155+ (%)	GC156 (%)	GC156+ (%)	GC157 (%)	GC157+ (%)	GC158 (%)	GC158+ (%)	GC159 (%)	GC159+ (%)	GC160 (%)	GC160+ (%)	GC161 (%)	GC161+ (%)	GC162 (%)	GC162+ (%)	GC163 (%)	GC163+ (%)	GC164 (%)	GC164+ (%)	GC165 (%)	GC165+ (%)	GC166 (%)	GC166+ (%)	GC167 (%)	GC167+ (%)	GC168 (%)	GC168+ (%)	GC169 (%)	GC169+ (%)	GC170 (%)	GC170+ (%)	GC171 (%)	GC171+ (%)	GC172 (%)	GC172+ (%)	GC173 (%)	GC173+ (%)	GC174 (%)	GC174+ (%)	GC175 (%)	GC175+ (%)	GC176 (%)	GC176+ (%)	GC177 (%)	GC177+ (%)	GC178 (%)	GC178+ (%)	GC179 (%)	GC179+ (%)	GC180 (%)	GC180+ (%)	GC181 (%)	GC181+ (%)	GC182 (%)	GC182+ (%)	GC183 (%)	GC183+ (%)	GC184 (%)	GC184+ (%)	GC185 (%)	GC185+ (%)	GC186 (%)	GC186+ (%)	GC187 (%)	GC187+ (%)	GC188 (%)	GC188+ (%)	GC189 (%)	GC189+ (%)	GC190 (%)	GC190+ (%)	GC191 (%)	GC191+ (%)	GC192 (%)	GC192+ (%)	GC193 (%)	GC193+ (%)	GC194 (%)	GC194+ (%)	GC195 (%)	GC195+ (%)	GC196 (%)	GC196+ (%)	GC197 (%)	GC197+ (%)	GC198 (%)	GC198+ (%)	GC199 (%)	GC199+ (%)	GC200 (%)	GC200+ (%)	GC201 (%)	GC201+ (%)	GC202 (%)	GC202+ (%)	GC203 (%)	GC203+ (%)	GC204 (%)	GC204+
--	-----	------------	----------	------	------------	-------------	--------	---------	----------	---------	----------	---------	----------	---------	----------	---------	----------	---------	----------	---------	----------	----------	-----------	----------	-----------	----------	-----------	----------	-----------	----------	-----------	----------	-----------	----------	-----------	----------	-----------	----------	-----------	----------	-----------	----------	-----------	----------	-----------	----------	-----------	----------	-----------	----------	-----------	----------	-----------	----------	-----------	----------	-----------	----------	-----------	----------	-----------	----------	-----------	----------	-----------	----------	-----------	----------	-----------	----------	-----------	----------	-----------	----------	-----------	----------	-----------	----------	-----------	----------	-----------	----------	-----------	----------	-----------	----------	-----------	----------	-----------	----------	-----------	----------	-----------	----------	-----------	----------	-----------	----------	-----------	----------	-----------	----------	-----------	----------	-----------	----------	-----------	----------	-----------	----------	-----------	----------	-----------	----------	-----------	----------	-----------	----------	-----------	----------	-----------	----------	-----------	----------	-----------	----------	-----------	----------	-----------	----------	-----------	----------	-----------	----------	-----------	----------	-----------	----------	-----------	----------	-----------	----------	-----------	----------	-----------	----------	-----------	----------	-----------	----------	-----------	----------	-----------	----------	-----------	----------	-----------	----------	-----------	----------	-----------	----------	-----------	----------	-----------	----------	-----------	----------	-----------	----------	-----------	----------	-----------	----------	-----------	----------	-----------	----------	-----------	----------	-----------	----------	-----------	----------	-----------	----------	-----------	----------	-----------	----------	-----------	----------	-----------	----------	-----------	----------	-----------	----------	-----------	----------	-----------	-----------	------------	-----------	------------	-----------	------------	-----------	------------	-----------	------------	-----------	------------	-----------	------------	-----------	------------	-----------	------------	-----------	------------	-----------	------------	-----------	------------	-----------	------------	-----------	------------	-----------	------------	-----------	------------	-----------	------------	-----------	------------	-----------	------------	-----------	------------	-----------	------------	-----------	------------	-----------	------------	-----------	------------	-----------	------------	-----------	------------	-----------	------------	-----------	------------	-----------	------------	-----------	------------	-----------	------------	-----------	------------	-----------	------------	-----------	------------	-----------	------------	-----------	------------	-----------	------------	-----------	------------	-----------	------------	-----------	------------	-----------	------------	-----------	------------	-----------	------------	-----------	------------	-----------	------------	-----------	------------	-----------	------------	-----------	------------	-----------	------------	-----------	------------	-----------	------------	-----------	------------	-----------	------------	-----------	------------	-----------	------------	-----------	------------	-----------	------------	-----------	------------	-----------	------------	-----------	------------	-----------	------------	-----------	------------	-----------	------------	-----------	------------	-----------	------------	-----------	------------	-----------	------------	-----------	------------	-----------	------------	-----------	------------	-----------	------------	-----------	------------	-----------	------------	-----------	------------	-----------	------------	-----------	------------	-----------	------------	-----------	------------	-----------	------------	-----------	------------	-----------	------------	-----------	------------	-----------	------------	-----------	------------	-----------	------------	-----------	------------	-----------	------------	-----------	------------	-----------	------------	-----------	------------	-----------	------------	-----------	------------	-----------	------------	-----------	------------	-----------	------------	-----------	------------	-----------	------------	-----------	------------	-----------	------------	-----------	------------	-----------	------------	-----------	------------	-----------	------------	-----------	------------	-----------	--------

DEFINITION	Homo sapiens EWSR1 full length open reading frame (ORF) cDNA clone (cDNA clone C220RF.pGEM.EWSR1).
ACCESSION	CR456490
VERSION	CR456490.1 GI:47678510
KEYWORDS	cDNA; chromosome 22; ORF.
SOURCE	Homo sapiens (human)
ORGANISM	Homo sapiens
TITLE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 2189)
REFERENCE	Collins,J.E., Wright,C.L., Edwards,C.A., Davis,M.P., Grinham,J.A., Cole,C.G., Goward,M.E., Aguado,B., Mallia,M., Mokrab,Y., Huckle,E.J., Beare,D.M. and Dunham,I. Direct Submission Submitted (24-May-2004) Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: c22g@sanger.ac.uk Manuscript Sanger Institute name : pGEM_EWSR1
JOURNAL	Homo sapiens cDNA sequence. This sequence was generated as part of The Wellcome Trust Sanger Institute program to isolate cDNA clones representing the full length open reading frame of well annotated protein coding genes on human chromosome 22. For more information see http://www.sanger.ac.uk/HGP/Chr22/. location/Qualifiers
COMMENT	
FEATURES	
source	1..2189 /organism="Homo sapiens" /mol_type="mRNA" /db_xref="taxon:9606" /chromosome="22" /clone="pGEM_EWSR1" /lab_host="JM109" 1..2189 /gene="EWSR1" 34..2001 /genome="EWSR1" /codon_start=1 /protein_id="CAG30376.1" /db_xref="GI:47678511"
gene	/translation="MASTDYSTVYSQAAGQGSAYTAQTPTGTATTAQTAAVGGVGYT GOPDVSATOAOTATATVGTATVATSYSGQPPTGYTPAPAPASPVGVSGTGAADITTT ATVTTOASVAAOSAVGTOPAPAYGQGPAAATAATRPDGNKPKTERISPOSSTGVNQQ PSLGGGNSYSPDPGVSYPMPATPAPSPPTPSYSTOPTSYDSSSVQNTYGOPS SYGQSSSYGQSSSYGQPPTSYPTQTSYGAQSQSSSYGQSSFPDDHSNG VYGDSGFSGFGENRMSGSDPNRKRGDFDRGMSRGGRGGMGAGERGVNKKT PGGMDEGPDLDLPVPDPEBDSNSAIYYGLDNSVTLDDLADFPGKAVVKNKKTK GQPMIHIYLDKETGKPKGDATVSEDEPTAKAAVEMWDGDFOSKLXVLARKPPMT NSMRGCP.PREGRMPEPLRGKPGCGPGKPGMGKGRGDGSGPRPARGSRGNP SGGNNVNRAGMDQCNPJGCGNMFEARTCCNOCKAPKRGFLPPPPPCGGRGREGS PGKRGRGGLMDRGPGGCMFRGGRGDRGDFRGKAMDGRFGGGRGRRGPGPPPL MEWGGRRGRGGRGKMDKEHRQERRDRPY"
CDS	
ORIGIN	
Query Match	99.1%; Score 2066; DB 9; Length 2189;
Best Local Similarity	99.9%; Pred. No. 0;
Matches 2019; Conservative	0; Mismatches 0; Indels 3; Gaps 1;
Oy	4 GAAGCTTGAGAAGAACGAGAGAGAGAAAATGCGTCACGAGATTACAGTACTAT 63
Dd	1 GGAGCTTGAGAAGAACGAGAGAGAGAAAATGCGTCACGAGATTACAGTACTAT 60
Oy	64 AGCCAAGCTGACGCGCAGCAGAGGCTACAGTGCCTTACACGCCGCCAAGCCCACTCAAAGATAT 123
Dd	61 AGCCAAGCTGACGCGCAGCAGAGGCTACAGTGCCTTACACGCCGCCAAGCCCACTCAAAGATAT 120
Oy	124 GCACAGACCACCCGAGGAGTATGGCCACAACAAAGCTATGGAACCTATGGACAGCCCATGAT 183
Dd	121 GCACAGACCACCCGAGGAGTATGGCCACAACAAAGCTATGGAACCTATGGACAGCCCATGAT 180
Oy	184 GTACAGCTATACCAAGGCTCAGACCACTGCAACCTATGGAGAGACCGGCTATGCACTTCT 243
Dd	181 GTACAGCTATACCAAGGCTCAGACCACTGCAACCTATGGAGAGACCGGCTATGCACTTCT 240
Oy	244 TATGAGACGCTCCCACTGGTTATATCTTCCACTGCCCCCCAGGAGATACAGCACGCT 303

Db 241 TATGACAGGCTCCCACTGGTTATACACTGCACTGCCCCGAGGATACAGCCAGCT 300  
 QY 304 GTCCAGGGGTATGACACTGGTCTTATGATACCACTGCTCAAGTCAACCAACCCAG 363  
 Db 301 GTCCAGGGGTATGACACTGGTCTTATGATACCACTGCTCAAGTCAACCAACCCAG 360  
 QY 364 GCGCTCCATGACGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 423  
 Db 361 GCGCTCCATGACGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 420  
 QY 424 CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 483  
 Db 421 CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 480  
 QY 484 CAACCTCAATCTGACACAGGGGGTTTACACACAGCCAGCCTAGAGATATGACAGAGTAC 543  
 Db 481 CAACCTCAATCTGACACAGGGGGTTTACACACAGCCAGCCTAGAGATATGACAGAGTAC 540  
 QY 544 TACGATATCCCAAGTACCTGGAGGCTACCCATGACAGCAGTCACTGACCTCCATCC 603  
 Db 541 TACGATATCCCAAGTACCTGGAGGCTACCCATGACAGCAGTCACTGACCTCCATCC 600  
 QY 604 TACCTCTCAACAGCTATCTCTCAACAGCCAGCCTAGTATGATCAGACAGTACTCT 663  
 Db 601 TACCTCTCAACAGCTATCTCTCAACAGCCAGCCTAGTATGATCAGACAGTACTCT 660  
 QY 664 CAGCAGAACTATGAGGCAACCGAGAGCTATGACAGCAGAGTACTATGCTCAACAA 723  
 Db 661 CAGCAGAACTATGAGGCAACCGAGAGCTATGACAGCAGAGTACTATGCTCAACAA 720  
 QY 724 AGCAGCTATGGGGGAGGAGGCTCCCACTAGTTACCCACCCCAACTGAGCTCAAGCAG 783  
 Db 721 AGCAGCTATGGGGGAGGAGGCTCCCACTAGTTACCCACCCCAACTGAGCTCAAGCAG 780  
 QY 784 GCTCAAGTCAATATAGCCAAACAGAGCAGCCTACGCGCAGCAGAGTTATCCAGCAG 843  
 Db 781 GCTCAAGTCAATATAGCCAAACAGAGCAGCCTACGCGCAGCAGAGTTATCCAGCAG 840  
 QY 844 GACCAACCCAGTACATGAGTGTGTTTATGAGGAGAGTCTGAGAGTTTTCGACACAGA 903  
 Db 841 GACCAACCCAGTACATGAGTGTGTTTATGAGGAGAGTCTGAGAGTTTTCGACACAGA 900  
 QY 904 GAGAAACCGGAGCAGATGAGTGGCTGATTAACCGGGGAGGAGGAGGAGGAGTTGATGCT 963  
 Db 901 GAGAAACCGGAGCAGATGAGTGGCTGATTAACCGGGGAGGAGGAGGAGGAGTTGATGCT 960  
 QY 964 GAGAGCAGTACAGAGTGGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1023  
 Db 961 GAGAGCAGTACAGAGTGGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1017  
 QY 1024 CGAGGTGGCTTCAATTAAGCTGTGAGACCCATGAGTGAAGACCAAGTCTTGAATGAGC 1083  
 Db 1018 CGAGGTGGCTTCAATTAAGCTGTGAGACCCATGAGTGAAGACCAAGTCTTGAATGAGC 1077  
 QY 1084 CCACCTGATGATCCAGATGAGACCTGAGCAACAGTGAATTTTATGATCAAGATTAAT 1143  
 Db 1078 CCACCTGATGATCCAGATGAGACCTGAGCAACAGTGAATTTTATGATCAAGATTAAT 1137  
 QY 1144 GACAGTGTGATCTAGATGATCTGAGCAGACTTCTTAAAGCAGTGTGGGTGTTAAAGTG 1203  
 Db 1138 GACAGTGTGATCTAGATGATCTGAGCAGACTTCTTAAAGCAGTGTGGGTGTTAAAGTG 1197  
 QY 1204 AACAGAGAACTGGGCAACCCATGATCAATCTTACCTGACAAAGGAAACGAGAAAGCCC 1263  
 Db 1198 AACAGAGAACTGGGCAACCCATGATCAATCTTACCTGACAAAGGAAACGAGAAAGCCC 1257  
 QY 1264 AAAGCGGATGCGACAGTGTCTTATGAAAGCCACCCATGCGCAAGGCGCGTGAATGG 1323  
 Db 1258 AAAGCGGATGCGACAGTGTCTTATGAAAGCCACCCATGCGCAAGGCGCGTGAATGG 1317  
 QY 1324 TTGTATGAGGAAGATTTTCAAGGAGCAAACTTAAAGTCTCCCTTGTCTGGAAGAGGCT 1383  
 Db 1318 TTGTATGAGGAAGATTTTCAAGGAGCAAACTTAAAGTCTCCCTTGTCTGGAAGAGGCT 1377

QY 1384 CCAATGAAACAGTATGGGGGGTGTCTGCCACCCCGTGAAGGGAGAGGATGCAACCA 1443  
 Db 1378 CCAATGAAACAGTATGGGGGGTGTCTGCCACCCCGTGAAGGGAGAGGATGCAACCA 1437  
 QY 1444 CTCCTGAGAGTCCAGAGAGCCAGAGAGTCTTGGGGGAGCCATGAGTGTGATGAGAGC 1503  
 Db 1438 CTCCTGAGAGTCCAGAGAGCCAGAGAGTCTTGGGGGAGCCATGAGTGTGATGAGAGC 1497  
 QY 1504 CCGTGAAGAGTATGAGAGAGGCTTCTCTCAAGAGACCCCGGGGTTCCGAGGAAACCC 1563  
 Db 1498 CCGTGAAGAGTATGAGAGAGGCTTCTCTCAAGAGACCCCGGGGTTCCGAGGAAACCC 1557  
 QY 1564 TCTGAGAGAGAAAGCTCCAGACCCAGAGTGTGAGACTGAGAGTGTGATGAGAGTGTG 1623  
 Db 1558 TCTGAGAGAGAAAGCTCCAGACCCAGAGTGTGAGACTGAGAGTGTGATGAGAGTGTG 1617  
 QY 1624 GGAACACAGAACTTCCCTGAGAAACAGAGTGCACCAAGTGTAAAGGCCCAAGCTGAA 1683  
 Db 1618 GGAACACAGAACTTCCCTGAGAAACAGAGTGCACCAAGTGTAAAGGCCCAAGCTGAA 1677  
 QY 1684 GCGTTCTCTCCCGCACCTTTTCGCGCCCGGGTGTGATGTGTGAGAGAGTGTGCTGT 1743  
 Db 1678 GCGTTCTCTCCCGCACCTTTTCGCGCCCGGGTGTGATGTGTGAGAGAGTGTGCTGT 1737  
 QY 1744 GCGATGCGGGGAGAGAGAGTGTGCTCATGATGTGTGTGTGCTCCGAGTGTGATGAG 1803  
 Db 1738 GCGATGCGGGGAGAGAGAGTGTGCTCATGATGTGTGTGTGCTCCGAGTGTGATGAG 1797  
 QY 1804 GGTGCGCGTGTGTGAGACAGAGTGTGCTTCCGTGTGTGCGCGGAGATGACCGAGTGTG 1863  
 Db 1798 GGTGCGCGTGTGTGAGACAGAGTGTGCTTCCGTGTGTGCGCGGAGATGACCGAGTGTG 1857  
 QY 1864 TTTGTGTGAGAGAAAGACAGAGTGTGCTTGGGGGGGCTTGGAGCTTGTGATGAAACAGAT 1923  
 Db 1858 TTTGTGTGAGAGAAAGACAGAGTGTGCTTGGGGGGGCTTGGAGCTTGTGATGAAACAGAT 1917  
 QY 1924 GAGAGAAAGAGAGAGAGAGTGTGAGACCTGGAATAAGATTAAGGCGAGCAGCTGAG 1983  
 Db 1918 GAGAGAAAGAGAGAGAGAGTGTGAGACCTGGAATAAGATTAAGGCGAGCAGCTGAG 1977  
 QY 1984 GAGCGAGAGATGAGGCTTACTAGATGACAGAGCCCGCAGAG 2025  
 Db 1978 GAGCGAGAGATGAGGCTTACTAGATGACAGAGCCCGCAGAG 2019

RESULT 8  
 BC072442  
 LOCUS  
 DEFINITION  
 Homo sapiens Ewing sarcoma breakpoint region 1, transcript variant  
 EMS, mRNA (CDNA clone MGC:87992 IMAGE:6047805), complete cds.  
 BC072442  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Homo sapiens (human)  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Hominidae; Homo.  
 1 (bases 1 to 2164)  
 Klausner,R.D., Collins,F.S., Wagner,L., Shennan,C.M., Schuler,G.D.,  
 Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Hahet,N.K.,  
 Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hahet,F.,  
 Diatchenko,J., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,  
 Schaefer,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,  
 Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,  
 Abramson,R.D., Mulhaly,S.J., Bosak,S.A., McKwan,P.J.,  
 McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,  
 Wotley,K.C., Hale,S., Garcia,A.M., Gay,L., Hulyk,S.W.,  
 Villalón,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,  
 Fahey,J., Helton,E., Kettelman,M., Madan,A., Rodriguez,S.,  
 Sanchez,A., Whitting,M., Madan,A., Young,A.C., Shevchenko,Y.,



Db 961 GGAGAGCGCGTGGAGTGG---GGCGTGGAGAGAGAGTGGCTTCAATAGCCTGGTGA 1017  
 Qy 1051 CCCATGATGAGAGACCAATCTTATGATCTAGGCCCACTTATGATCCAGATGAAGTCT 1110  
 Db 1018 CCCATGATGAGAGACCAATCTTATGATCTAGGCCCACTTATGATCCAGATGAAGTCT 1077  
 Qy 1111 GACAAAGTGTCAATTTATGTAACAAGATTAAAGAGTGTGACTGTAGATGATCTGGCA 1170  
 Db 1078 GACAAAGTGTCAATTTATGTAACAAGATTAAAGAGTGTGACTGTAGATGATCTGGCA 1137  
 Qy 1171 GACTTTTAAAGAGTGTGGGTTGTTAAGATGAAACAAGAACTGGGCAACCATGATC 1230  
 Db 1138 GACTTTTAAAGAGTGTGGGTTGTTAAGATGAAACAAGAACTGGGCAACCATGATC 1197  
 Qy 1231 CACATCTACTGGAACAAGAAACAAGAAAGCCCAAGGCGATGCCACAGTGTCTATGAA 1290  
 Db 1198 CACATCTACTGGAACAAGAAACAAGAAAGCCCAAGGCGATGCCACAGTGTCTATGAA 1257  
 Qy 1291 GACCCACCCCACTGCGAAGGCTGCGGTGGAAATGTTGATGGGAAAGATTTCAGAGGAGC 1350  
 Db 1258 GACCCACCCCACTGCGAAGGCTGCGGTGGAAATGTTGATGGGAAAGATTTCAGAGGAGC 1317  
 Qy 1351 AAACTTAAAGTCTCCCTGCTGGAGAGAGCCTTCAATGAAACAGTATCGGGTGTCTG 1410  
 Db 1318 AAACTTAAAGTCTCCCTGCTGGAGAGAGCCTTCAATGAAACAGTATCGGGTGTCTG 1377  
 Qy 1411 CCAACCCCTGAGAGGCGAGAGCATGCAACCACTCCGTGAGAGTCCAGAGGCCAGGA 1470  
 Db 1378 CCAACCCCTGAGAGGCGAGAGCATGCAACCACTCCGTGAGAGTCCAGAGGCCAGGA 1437  
 Qy 1471 GGTCCTGGGGGAGCCCATGGGTGCGATGGGAGGCGCGTGGAGAGATAGAGAGGCTTCCCT 1530  
 Db 1438 GGTCCTGGGGGAGCCCATGGGTGCGATGGGAGGCGCGTGGAGAGATAGAGAGGCTTCCCT 1497  
 Qy 1531 CCAAGAGAGACCCCGGGGTTCCCGAGAGAGACCCCTCTGAGAGAGAGAACTCCAGACCGA 1590  
 Db 1498 CCAAGAGAGACCCCGGGGTTCCCGAGAGAGACCCCTCTGAGAGAGAGAACTCCAGACCGA 1557  
 Qy 1591 GCTGAGAGACTGGCATGTCTCCCAATCCGGGTTGTGAGAAACAAGAACTTCCCTGAGAGAA 1650  
 Db 1558 GCTGAGAGACTGGCATGTCTCCCAATCCGGGTTGTGAGAAACAAGAACTTCCCTGAGAGAA 1617  
 Qy 1651 GAGTGCAACAGTGTAGAGGCCCAAGAGGCTTCCCTCCCGGCAACCTTCCGCGCC 1710  
 Db 1618 GAGTGCAACAGTGTAGAGGCCCAAGAGGCTTCCCTCCCGGCAACCTTCCGCGCC 1677  
 Qy 1711 CCGGCTGTGATGCTGTCAGAGGTGCGCTGTGTGTCATGCGGGAGAGAAAGGTGCGCTC 1770  
 Db 1678 CCGGCTGTGATGCTGTCAGAGGTGCGCTGTGTGTCATGCGGGAGAGAAAGGTGCGCTC 1737  
 Qy 1771 ATGATGCTGTGCTGCTCCCGTGGAAATTTCAAGAGTGGCCGTGTGAGACAAGAGTGGC 1830  
 Db 1738 ATGATGCTGTGCTGCTCCCGTGGAAATTTCAAGAGTGGCCGTGTGAGACAAGAGTGGC 1797  
 Qy 1831 TTCCTGTGTGGCGGGGATGAGACCGAGAGTGGCTTGTGTGAGAGAAACAAGTGGGCT 1890  
 Db 1798 TTCCTGTGTGGCGGGGATGAGACCGAGAGTGGCTTGTGTGAGAGAAACAAGTGGGCT 1857  
 Qy 1891 GGGGGGGCCCTTGAACCTTTGATGAGAAACAAGTGGAGAAAGAGAGAGAGAGTGGAGGA 1950  
 Db 1858 GGGGGGGCCCTTGAACCTTTGATGAGAAACAAGTGGAGAAAGAGAGAGAGAGTGGAGGA 1917  
 Qy 1951 CTTGAGAAAATGATTAAGGAGACCGCTGAGAGCGCAGAGATCGGCCCTTACTAGATG 2010  
 Db 1918 CTTGAGAAAATGATTAAGGAGACCGCTGAGAGCGCAGAGATCGGCCCTTACTAGATG 1977  
 Qy 2011 CAGAGACCCCGCAGGA 2025  
 Db 1978 CAGAGACCCCGCAGGA 1992

RESULT 9

BC011048  
 LOCUS 2364 bp mRNA linear PRI 29-JUN-2004  
 DEFINITION Homo sapiens Ewing sarcoma breakpoint region 1, transcript variant  
 EMS, mRNA (CDNA clone MGC:16900 IMAGE:4340116), complete cds.  
 ACCESSION BC011048  
 VERSION BC011048.1 GI:15029674  
 KEYWORDS MGC.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE  
 1 (bases 1 to 2364)  
 Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G.,  
 Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D.,  
 Altschul, S.F., Zeeberg, B., Buco, K.H., Scheffer, C.F., Bhat, N.K.,  
 Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, P.,  
 Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,  
 Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L.,  
 Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S.,  
 Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J.,  
 Abramson, R.D., Mullahy, S.J., Bosak, S.A., McEwan, P.J.,  
 McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S.,  
 Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W.,  
 Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,  
 Fahey, J., Helton, E., Kettner, M., Madan, A., Rodriguez, S.,  
 Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shenchenko, Y.,  
 Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D.,  
 Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,  
 Butlerfield, Y.S., Krzywinski, M.I., Skliska, U., Smal, D.E.,  
 Scherch, A., Schein, J.E., Jones, S.J., and Marra, M.A.  
 Generation and initial analysis of more than 15,000 full-length  
 human and mouse cDNA sequences  
 Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16699-16903 (2002)  
 12477932  
 2 (bases 1 to 2364)  
 Strausberg, R.  
 Direct Submissions  
 Submitted (25-JUN-2001) National Institutes of Health, Mammalian  
 Gene Collection (MGC), Cancer Genomics Office, National Cancer  
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2550,  
 USA  
 NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
 Contact: MGC help desk  
 Email: [cgabs-remail.nih.gov](mailto:cgabs-remail.nih.gov)  
 Tissue Procurement: Lou Staudt  
 CDNA Library Preparation: Life Technologies, Inc.  
 DNA Sequencing by: Baylor College of Medicine Human Genome  
 Sequencing Center  
 Center code: BCM-HGSC  
 Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>  
 Contact: [amg@bcm.tmc.edu](mailto:amg@bcm.tmc.edu)  
 Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Lounsbeg, H.,  
 Kowls, C.R., Sneed, A.J., Martin, R.G., Muzny, D.M., Nanavati,  
 A.N., Gibbs, R.A.  
 Clone distribution: MGC clone distribution information can be found  
 through the I.M.A.G.E. Consortium/LLNL ac: <http://image.llnl.gov>  
 Series: IRAX Plate: 12 Row: P Column: 1  
 This clone was selected for full length sequencing because it  
 passed the following selection criteria: matched mRNA gi: 4885224.  
 Location/Qualifiers  
 1..2364  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="MGC:16900 IMAGE:4340116"  
 /tissue\_type="Lymph, Lymphoma"  
 /clone\_lib="NIH\_MGC\_95"  
 /lab\_host="DH10B"  
 /note="Vector: PCMV-SPORT6"  
 1..2364  
 /gene="EWSR1"

gene





Db	1761	GGAATGCTGTGTGTCGCCGGTGGAAAGTTTCAGAGGTGTCCGTGTGTGAGACAGAGGTGGCTT	1820
Oy	1833	CCGTGTGTGCGCGGGCATGACCCGAGTGCTGTGTGTGTGAGGAGACGAGGTGGCCTTGG	1892
Db	1821	CCGTGTGTGCGCGGGCATGAGCACGAGGTGTGCTTTGTGTGAGGAGACGAGGTGGCCTTGG	1880
Oy	1893	GGGGCCCCCTTGAGCCTTTGATTGTGAACAAGATGGGAGAAAGAAGAGAGAGAGCGTGAAGACC	1952
Db	1881	GGGGCCCCCTTGAGCCTTTGATTGTGAACAAGATGGGAGAAAGAAGAGAGAGAGCGTGAAGACC	1940
Oy	1953	TGAAAAAATGATAAAGCGCAGCAACCTTCAGAGCGCAGAGATGCCTTAAGTAGCA	2012
Db	1941	TGAAAAAATGATAAAGCGCAGCAACCTTCAGAGCGCAGAGATGCCTTAAGTAGCA	2000
Oy	2013	GAGACCCCGCAGCA	2025
Db	2001	GAGACCCCGCAGCA	2013
RESULT 10	Bt007796	1971 bp mRNA linear SYN 13-MAY-2003	
LOCUS	Bt007796	Synthetic construct Homo sapiens Ewing sarcoma breakpoint region 1	
DEFINITION		mRNA, partial cds.	
ACCESSION	Bt007796		
VERSION	Bt007796.1	GI:30584430	
KEYWORDS	FLI CDNA.		
SOURCE		synthetic construct	
ORGANISM		synthetic construct	
REFERENCE		other sequences; artificial sequences.	
AUTHORS		1 (bases 1 to 1971) Kaindl, N., Chen, X., Rolfs, A., Halleck, A., Hines, L., Eisenstein, S., Koundinya, M., Raphael, J., Moreira, D., Kelley, T., Labaer, J., Lin, Y., Phelan, M. and Farmer, A.	
TITLE		Cloning of human full-length CDSs in BD Creator (TM) System Donor vector	
JOURNAL		Unpublished	
REFERENCE		2 (bases 1 to 1971)	
AUTHORS		Kaindl, N., Chen, X., Rolfs, A., Halleck, A., Hines, L., Eisenstein, S., Koundinya, M., Raphael, J., Moreira, D., Kelley, T., Labaer, J., Lin, Y., Phelan, M. and Farmer, A.	
TITLE		Direct Submission	
JOURNAL		Submitted (13-MAY-2003) BD Biosciences Clontech, 1020 East Meadow Circle, Palo Alto, CA 94303, USA	
COMMENT		This CDS clone is a part of a collection of human full length expression clones generated by BD Biosciences Clontech and the Harvard Institute of Proteomics. Each CDS has been cloned in two forms: with and without stop-codon (to allow fusion with C-terminal tag). The CDS has been directionally cloned using BD In-Fusion(TM) cloning system between the SalI and HindIII sites of the pDNR-DUAL vector. Additional sequences in the clone: "ACC" after SalI site and before "ATG" to provide Kozak consensus sequence; "GG" after last codon and before HindIII site to maintain reading frame. Clone distribution: http://biointo.clontech.com/orfclones.	
FEATURES		Location/Qualifiers	
source		1..1971	
		/organism="synthetic construct"	
		/mol_type="mRNA"	
		/db_xref="taxon:32630"	
		/clone="GH00297L1.0"	
		/clone_1lb="BD Creator (TM) CDS library derived from MGC collection"	
		/lab_host="DH5alpha TI resistant"	
		/note="Vector: pDNR-Dual1"	
		1..>1971	
		/note="Mutations: 1970:Stop->Leu"	
		/codon_start=1	
		/transl_table=11	
		/product="Homo sapiens Ewing sarcoma breakpoint region 1"	
		/protein_id="AAP3468.1"	
		/db_xref="gi:30584431"	
		/translation="WASTDYSTYISQAADQGYSAVTAPDTGTGACTQTQAVQQGSYGTVV	
CDS			

Query Match	97.3%: Score 1669.4; DB 12; Length 1971;	
Best Local Similarity 99.9%; Pred. No. 0;		
Matches 1970; Conservative 0; Mismatches 1; Indels 0; Gaps 0;		
QY 37 ATGGCGTCCACGAGATTACAGTACCTTATAGCCAGAGCTGCAGGCGAGCAGAGGGCTACAGTCT	96	
DB 1 ATGGCGTCCACGAGATTACAGTACCTTATAGCCAGAGCTGCAGGCGAGCAGAGGGCTACAGTCT	60	
QY 97 TACACCGGCCAGCCCACTCAAGGATATGCAACAGCCACCGAGCATATGGGCAACAAGC	156	
DB 61 TACACCGGCCAGCCCACTCAAGGATATGCAACAGCCACCGAGCATATGGGCAACAAGC	120	
QY 157 TATGGAACTATGAGCAGGCCCACTGATGTCAGTATCCAGGGCTCAGACCACTGCAACC	216	
DB 121 TATGGAACTATGAGCAGGCCCACTGATGTCAGTATCCAGGGCTCAGACCACTGCAACC	180	
QY 217 TATGGAGAGACCGGCTATGCACTTCTTATGAGAGGCTCCACCTGGTTATACTTCA	276	
DB 181 TATGGAGAGACCGGCTATGCACTTCTTATGAGAGGCTCCACCTGGTTATACTTCA	240	
QY 277 ACTGCCCCCAGGAGATACAGCCAGCTGTGCCAGGGATATGGCACTGGTCTTATGATACC	336	
DB 241 ACTGCCCCCAGGAGATACAGCCAGCTGTGCCAGGGATATGGCACTGGTCTTATGATACC	300	
QY 337 ACCACTCTACAGTACCAACCAACGAGGCTCCATGAGAGTCAAGTGTGCAATATGGCACT	396	
DB 301 ACCACTCTACAGTACCAACCAACGAGGCTCCATGAGAGTCAAGTGTGCAATATGGCACT	360	
QY 397 CAGGCTCTTATTCAGGCTATGAGGACAGCAGCAGCACTGCACTTACAAGACCGAG	456	
DB 361 CAGGCTCTTATTCAGGCTATGAGGACAGCAGCAGCACTGCACTTACAAGACCGAG	420	
QY 457 GATGAAAACAAGCCCACTGAGACTGATGCACTCAATCTTACACAAGGGGGTTACAACAG	516	
DB 421 GATGAAAACAAGCCCACTGAGACTGATGCACTCAATCTTACACAAGGGGGTTACAACAG	480	
QY 517 CCCAGCTTAGGATATGAGCAGAGTAACTACAGTTATCCCAAGGTATCCGGGAGCTACACCC	576	
DB 481 CCCAGCTTAGGATATGAGCAGAGTAACTACAGTTATCCCAAGGTATCCGGGAGCTACACCC	540	
QY 577 ATGAGAGCAAGCATGCACTGCACCTCAATCCCTCTTACAGCACTATTCTCTTACAGACCG	636	
DB 541 ATGAGAGCAAGCATGCACTGCACCTCTTACAGCACTATTCTCTTACAGACCG	600	
QY 637 ACTAGTATGATACAGAGAGTTACTTTCAGGAAACCTATGGGCAACCGAGAGCTAT	696	
DB 601 ACTAGTATGATACAGAGAGTTACTTTCAGGAAACCTATGGGCAACCGAGAGCTAT	660	
QY 697 GGAACAGAGAGTACTATGGTCAACAAAGCAGCTATGGGACAGAGCTTCCCACTAGTTAC	756	
DB 661 GGAACAGAGAGTACTATGGTCAACAAAGCAGCTATGGGACAGAGCTTCCCACTAGTTAC	720	
QY 757 CCACCCCAAACTGGATCCTTACAGCAAGCTCCAGTCACTATAGCAACAGAGCAGAGC	816	
DB 721 CCACCCCAAACTGGATCCTTACAGCAAGCTCCAGTCACTATAGCAACAGAGCAGAGC	780	
QY 817 TACGGGACAGAGATTCAATCCGACAGGACCAACCCAGTAGACAGGAGTTATGGGAG	876	
DB 781 TACGGGACAGAGATTCAATCCGACAGGACCAACCCAGTAGACAGGAGTTATGGGAG	840	



QY 877 GAGCTGAGGAGATTTTCCGGAACGAGAGAGAACCGAGCATGAGTGGCCCTGATTAACCGG 936  
 DB 841 GAGCTGAGGAGATTTTCCGGAACGAGAGAGAACCGAGCATGAGTGGCCCTGATTAACCGG 900  
 QY 937 GGCAGGGGAAAGGGGGATTTGATCGTAGAGCAATGAGAGAGTGGCCGGAGAGAGA 996  
 DB 901 GGCAGGGGAAAGGGGGATTTGATCGTAGAGCAATGAGAGAGTGGCCGGAGAGAGA 960  
 QY 997 CGCGGTGGAATGGCAGCCCTGAGAGAGAGAGTGGCTTCAATTAAGCTGTGAGACCCATG 1056  
 DB 961 CGCGGTGGAATGGCAGCCCTGAGAGAGAGAGTGGCTTCAATTAAGCTGTGAGACCCATG 1020  
 QY 1057 GATGAAGGACCAATCTTGAATCTAGGCGCACTGTAGATCCAGATGAAGACTGCAAC 1116  
 DB 1021 GATGAAGGACCAATCTTGAATCTAGGCGCACTGTAGATCCAGATGAAGACTGCAAC 1080  
 QY 1117 AGTGCATTTATGATACAGAGATTAATGACAGTGTGACTAGATGATCTGAGACCTTC 1176  
 DB 1081 AGTGCATTTATGATACAGAGATTAATGACAGTGTGACTAGATGATCTGAGACCTTC 1140  
 QY 1177 TTTAAGAGTGTGGGTTTGAATGATGAAGAACTGGGCAACCCATGATCAATC 1236  
 DB 1141 TTTAAGAGTGTGGGTTTGAATGATGAAGAACTGGGCAACCCATGATCAATC 1200  
 QY 1237 TACCTGGACAAAGGAAACAGGAAAGCCGATGCCACAGTGTCTATGAAGACCA 1296  
 DB 1201 TACCTGGACAAAGGAAACAGGAAAGCCGATGCCACAGTGTCTATGAAGACCA 1260  
 QY 1297 CCCACTGCCAAGGCTGCGGTGGAATGTTGATGGAGAAATTTTCAAGGAGCAACTT 1356  
 DB 1261 CCCACTGCCAAGGCTGCGGTGGAATGTTGATGGAGAAATTTTCAAGGAGCAACTT 1320  
 QY 1357 AAGTGTCTCTTGTGCGAAGAAAGCTTCAATGAACGATGCGGGTGTGTTGCCACC 1416  
 DB 1321 AAGTGTCTCTTGTGCGAAGAAAGCTTCAATGAACGATGCGGGTGTGTTGCCACC 1380  
 QY 1417 CGTGAAGGCGAGGCAATGCCACCACTCCGTGAGAGTCCAGAGAGCCCGAGAGGTCCT 1476  
 DB 1381 CGTGAAGGCGAGGCAATGCCACCACTCCGTGAGAGTCCAGAGAGCCCGAGAGGTCCT 1440  
 QY 1477 GGGGAGCCCATGGGTGCGATGAGAGAGCGGTGAGAGATGAGAGAGGCTTCCCTCAAGA 1536  
 DB 1441 GGGGAGCCCATGGGTGCGATGAGAGAGCGGTGAGAGATGAGAGAGGCTTCCCTCAAGA 1500  
 QY 1537 GGAAGGCGGGGTTCCGAGGGAACCCCTCTGAGAGAGAAACGTCGACACCGAGCTGA 1596  
 DB 1501 GGAAGGCGGGGTTCCGAGGGAACCCCTCTGAGAGAGAAACGTCGACACCGAGCTGA 1560  
 QY 1597 GACTGGCAGTGTCCCAATCCGGGTGTGGAACCAAGAACTTCCGCTGAGAGACAGATGC 1656  
 DB 1561 GACTGGCAGTGTCCCAATCCGGGTGTGGAACCAAGAACTTCCGCTGAGAGACAGATGC 1620  
 QY 1657 AACCAAGTGAAGGCGCCCAAGGCTTCTCCGCGACACCTTTCGCGCCCGGGGT 1716  
 DB 1621 AACCAAGTGAAGGCGCCCAAGGCTTCTCCGCGACACCTTTCGCGCCCGGGGT 1680  
 QY 1717 GGTGATCTGAGCAAGAGTGGCCCTGATGAGATGCGGGAGAGAAAGAGTGGCTCATGAT 1776  
 DB 1681 GGTGATCTGAGCAAGAGTGGCCCTGATGAGATGCGGGAGAGAAAGAGTGGCTCATGAT 1740  
 QY 1777 CGTGTGTGTCCCGGTGGAATGTTCAAGAGTGGCCGTGTGAGACAGAGTGGCTTCGT 1836  
 DB 1741 CGTGTGTGTCCCGGTGGAATGTTCAAGAGTGGCCGTGTGAGACAGAGTGGCTTCGT 1800  
 QY 1837 GGTGGCCGGGGCAATGACCGAGGTTGTTGTGAGAGAAAGAGAGTGGCCCTGGGGGG 1896  
 DB 1801 GGTGGCCGGGGCAATGACCGAGGTTGTTGTGAGAGAAAGAGAGTGGCCCTGGGGGG 1860  
 QY 1897 CCCCCTGAGACTTTGATGAGAAACAGATGAGAGAGAGAAAGAGAGACTGAGAGACTGGA 1956  
 DB 1861 CCCCCTGAGACTTTGATGAGAAACAGATGAGAGAGAGAAAGAGAGACTGAGAGACTGGA 1920  
 QY 1957 AAAATGATTAAGGCGACACCGTCAGAGAGCCAGAGATGGCCCTTACTAG 2007.

Db 1921 AAAATGATTAAGGCGACACCGTCAGAGAGCCAGAGATGGCCCTTACTTG 1971

RESULT 11  
 AK056309  
 LOCUS  
 DEFINITION  
 Homo sapiens cDNA FLJ1747 f1s, clone NT2R12007377, highly similar to RNA-BINDING PROTEIN EWS.  
 AK056309  
 AK056309.1 GI:16551673  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM

# REFERENCE

Oca, T., Suzuki, Y., Nishikawa, T., Otsuki, T., Sugiyama, T., Irie, R., Makatsugu, A., Hayashi, K., Sato, H., Nagai, K., Kimura, K., Makita, H., Sekine, M., Odayashi, M., Nishi, T., Shibahara, T., Tanaka, F., Ishii, S., Yamamoto, J., Saito, K., Kawai, Y., Isono, Y., Nakamura, Y., Nagaharai, K., Murakami, K., Yaeuda, T., Iwayanagi, T., Wagatsuna, M., Shiratori, A., Sudo, H., Hosoiri, T., Kaku, Y., Kodaira, H., Kondo, H., Sugawara, M., Takahashi, M., Kanda, K., Yokoi, T., Furuya, T., Kikawa, E., Omura, Y., Abe, K., Kamihara, K., Katsura, N., Sato, K., Tanikawa, M., Yamazaki, M., Nimomiyama, K., Ishibashi, T., Yamashita, H., Hirakawa, K., Fujimori, K., Tanai, H., Kimata, M., Watanabe, S., Yoshida, S., Chiba, Y., Ishida, S., Oho, Y., Takiguchi, S., Watanabe, S., Hara, H., Tanase, T., Nomura, J., Togaya, S., Komai, F., Hara, R., Takeuchi, K., Arita, M., Imose, N., Kusano, J., Kamehori, K., Yuki, H., Ohnima, A., Sasaki, N., Aotsuka, S., Yoshikawa, Y., Matsunawa, H., Ichihara, T., Shionata, N., Sano, S., Moriya, S., Momiyama, H., Satoh, N., Takami, S., Terashima, Y., Suzuki, O., Nakagawa, S., Senoh, A., Mizoguchi, H., Goto, Y., Shimizu, F., Wakebe, H., Hishigaki, H., Watanabe, T., Sugiyama, A., Takemoto, M., Kawakami, B., Yamazaki, M., Watanabe, K., Kamegaki, A., Itakura, S., Fukuzumi, Y., Fujimori, Y., Komiyama, M., Tashiro, H., Taniguchi, A., Fujizawa, T., Oho, T., Yamada, K., Fujii, Y., Ozaki, K., Hiro, M., Omori, Y., Kawabata, A., Hiki, T., Kobatake, N., Inagaki, H., Ikema, Y., Okamoto, S., Okitani, R., Kawakami, T., Noguchi, S., Itoh, T., Shigeta, K., Senba, T., Matsunaga, K., Nakajima, Y., Mizuno, T., Morinaga, M., Sasaki, M., Togaishi, T., Oyama, M., Hata, H., Watanabe, M., Komatsu, T., Mizushima-Sugano, T., Nagase, T., Nomura, N., Kikuchi, H., Masuno, Y., Yamashita, R., Nakai, K., Yada, T., Nakamura, Y., Ohara, O., Isogai, T., and Sugano, S. Complete sequencing and characterization of 21,243 full-length human cDNAs Nat. Genet. 36 (1), 40-45 (2004) 14702039

2  
 Iehibashi, T., Kanehori, K., Yoshida, M., Watanabe, S., Ishida, S., Oho, Y., Houta, T., Hiraoka, S., Murakawa, K., Takiguchi, S., Kusano, T., Watanabe, M., Fujimori, K., Tanai, H., Iehida, M., Yamashita, H., Chiba, Y., Sugiyama, T., Irie, R., Otsuki, T., Sato, H., Makatsugu, A., Ishii, S., Yamamoto, J., Isono, Y., Kawai, H., O., Saito, K., Nishikawa, T., Kimura, K., Matsuno, K., Nakamura, Y., Sekine, M., Kikuchi, H., Kanda, K., Wagatsuna, M., Takahashi, F., Oshima, A., Sugiyama, A., Kawakami, B., Suzuki, Y., Sugano, S., Nagaharai, K., Masuno, Y., Nagai, K., and Isogai, T. NEDO human cDNA sequencing project Unpublished

3 (bases 1 to 2189)  
 Isogai, T., Otsuki, T., and Sugiyama, T. Submitted (24-OCT-2001) Takao Isogai, Helix Research Institute, Genomics Laboratory; 1532-3 Yama, Kisarazu, Chiba 252-0812, Japan (E-mail: genomics@helix.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986) NEDO human cDNA sequencing project supported by Ministry of Economy, Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology (RAB) (cDNA library construction: Helix Research Institute (HRI) (supported by Japan Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB,

## COMMENT



QY 1618 GGTGTGGAACACGAACTTGCTGAGAAACAGATGCAACAGTGTAAAGCCCAAG 1677  
DB 1618 GGTGTGGAACACGAACTTGCTGAGAAACAGATGCAACAGTGTAAAGCCCAAG 1677  
QY 1678 CTGGAAGCTTCTTCGCCCACTTTTCGCCGCCCGGGTGTATCTGTGCAAGGTGGC 1737  
DB 1678 CTGGAAGCTTCTTCGCCCACTTTTCGCCGCCCGGGTGTATCTGTGCAAGGTGGC 1737  
QY 1738 CTTGTGTGCAATGCGGGGAGGAAGAGTGGCTCATGTATGCTGTGTCTCCGGTGAATG 1797  
DB 1738 CTTGTGTGCAATGCGGGGAGGAAGAGTGGCTCATGTATGCTGTGTCTCCGGTGAATG 1797  
QY 1798 TTCAAGAGTGGCCCTGTGTGAGACAGAGTGGCTTCCGTGTGGCCCGGGCATGACGA 1857  
DB 1798 TTCAAGAGTGGCCCTGTGTGAGACAGAGTGGCTTCCGTGTGGCCCGGGCATGACGA 1857  
QY 1858 GGTGGCTTTGGTGTGAGAAAGACAGAGTGGCCCTGTGGGGGCCCTGTGACCTTTGATGAA 1917  
DB 1858 GGTGGCTTTGGTGTGAGAAAGACAGAGTGGCCCTGTGGGGGCCCTGTGACCTTTGATGAA 1917  
QY 1918 CAGATGGGAGGAAGAAGAGAGAGCCTGTGAGAACCTTGAAATAAGAGGACGAC 1977  
DB 1918 CAGATGGGAGGAAGAAGAGAGAGCCTGTGAGAACCTTGAAATAAGAGGACGAC 1977  
QY 1978 CGTCAGAGAGCAGAGATCGGCTTACTAGATGACAGACCCCGCAGA 2025  
DB 1978 CGTCAGAGAGCAGAGATCGGCTTACTAGATGACAGACCCCGCAGA 2025

RESULT 12  
LOCUS CQ730352 2326 bp DNA linear PAT 03-FEB-2004  
DEFINITION Sequence 16286 from Patent WO02068579.  
ACCESSION CQ730352 GI:42303963  
VERSION CQ730352.1  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE  
AUTHORS Venter, C.J., Adams, M.C., Li, P.W. and Myers, E.W.  
TITLE RfEs, such as nucleic acid arrays, comprising a majority of  
humanexons or transcripts, for detecting expression and other uses  
theroeof

JOURNAL Patent: WO 02068579-A 16286 06-SEP-2002;  
PE Corporation (NY) (US)

FEATURES  
source location/Qualifiers  
1..2326

/organism="Homo sapiens"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:9606"

## ORIGIN

Query Match 97.0%; Score 1964.4; DB 6; Length 2326;  
Best Local Similarity 99.9%; Pred. No. 0;  
Matches 1976; Conservative 0; Mismatches 1; Indels 1; Gaps 1;  
QY 49 GATTACAGTACTATAGCAAGCTGACGACGAGGGCTACAGTGTTCACACGCGCCAG 108  
DB 8 GATTACAGTACTATAGCAAGCTGACGACGAGGGCTACAGTGTTCACACGCGCCAG 67  
QY 109 CCCACTCAGAGATATGACACAGACCCAGGCAATATGGGCAACAAAGCTATGAACTAT 168  
DB 68 CCCACTCAGAGATATGACACAGACCCAGGCAATATGGGCAACAAAGCTATGAACTAT 127  
QY 169 GACAGGCGCACTGATGCTATACCCAGGCTCAGACCTGGAACCTATGAGGACGAC 228  
DB 128 GACAGGCGCACTGATGCTATACCCAGGCTCAGACCTGGAACCTATGAGGACGAC 187  
QY 229 GCTGATGCACTTCTTATGACAGCCTCCACCTGTTATCTACTCCAACTGCCCCAG 288  
DB 188 GCTGATGCACTTCTTATGACAGCCTCCACCTGTTATCTACTCCAACTGCCCCAG 247

QY 289 GCAATACAGCAGCCTGTGCAAGGGGTATGGAATGAGCTTATGATACCAACACTGCTACA 348  
DB 248 GCAATACAGCAGCCTGTGCAAGGGGTATGGAATGAGCTTATGATACCAACACTGCTACA 307  
QY 349 GTCAACACCAACCGAGCCTCTTATGACACTCACTGTCATATGACACTGACCTGCTTAT 408  
DB 308 GTCAACACCAACCGAGCCTCTTATGACACTCACTGTCATATGACACTGACCTGCTTAT 367  
QY 409 CCAAGCTTATGAGCAGACGACGACCACTGACCTTACAAAGCCGACAGATGGAACAG 468  
DB 368 CCAAGCTTATGAGCAGACGACGACCACTGACCTTACAAAGCCGACAGATGGAACAG 427  
QY 469 CCCACTGAGACTGTACCACTCAATCTAGACACAGGGGGTTTACAAACGACCCACTGAG 528  
DB 428 CCCACTGAGACTGTACCACTCAATCTAGACACAGGGGGTTTACAAACGACCCACTGAG 487  
QY 529 TATGACAGAGTAACTACATTAATCCCAAGGTACTGGAAGCTACCCCATGACGACGTC 588  
DB 488 TATGACAGAGTAACTACATTAATCCCAAGGTACTGGAAGCTACCCCATGACGACGTC 547  
QY 589 ACTGACCTTCATCTTACCTCTTACCAAGCTATTCCTTACACAGCCGACTAGTTATGAT 648  
DB 548 ACTGACCTTCATCTTACCTCTTACCAAGCTATTCCTTACACAGCCGACTAGTTATGAT 607  
QY 649 CAGAGCAGTTACTCTACAGCAACACCTATGGGCAACCGACAGCTATGACAGACAGAT 708  
DB 608 CAGAGCAGTTACTCTACAGCAACACCTATGGGCAACCGACAGCTATGACAGACAGAT 667  
QY 709 AGCTATGATCAACAAAGCAGCTATGGGACAGCAGCTCCACTGTATCCACCCCAACT 768  
DB 668 AGCTATGATCAACAAAGCAGCTATGGGACAGCAGCTCCACTGTATCCACCCCAACT 727  
QY 769 GATCTCTACAGCAAGCTCCAGTCAATATAGCAACAGACAGCACTTACGAGCAG 828  
DB 728 GATCTCTACAGCAAGCTCCAGTCAATATAGCAACAGACAGCACTTACGAGCAG 787  
QY 829 AGTTCAATTCGACAGCAACCCCACTAGTGGTGTATGAGCAGAGTCTGAGGA 888  
DB 788 AGTTCAATTCGACAGCAACCCCACTAGTGGTGTATGAGCAGAGTCTGAGGA 847  
QY 889 TTTTCCGGACACAGAGAAACGGAGCATGAGTGGCTGTATTAACGGGGACGGGGAGA 948  
DB 848 TTTTCCGGACACAGAGAAACGGAGCATGAGTGGCTGTATTAACGGGGACGGGGAGA 907  
QY 949 GGGGATTTGATGCTGAGAGCAGATGAGCAGAGTGGGCGGGAGAGAGACCGGTGATG 1008  
DB 908 GGGGATTTGATGCTGAGAGCAGATGAGCAGAGTGGGCGGGAGAGAGACCGGTGATG 967  
QY 1009 GGCAGGCTGAGAGAGGAGGTGCTTCAATATAGCTGTGTGACCCATGATGAAGCA 1068  
DB 968 GGCAGGCTGAGAGAGGAGGTGCTTCAATATAGCTGTGTGACCCATGATGAAGCA 1027  
QY 1069 GATCTGATCTAGCCCACTGATGATCCAGATGAAGCACTGACCAACAGTGAATTTAT 1128  
DB 1028 GATCTGATCTAGCCCACTGATGATCCAGATGAAGCACTGACCAACAGTGAATTTAT 1087  
QY 1129 GTACAGAGTTAATATGACAGTGTGACTTATGATCTGGCAACTTCTTAAAGAGTGT 1188  
DB 1088 GTACAGAGTTAATATGACAGTGTGACTTATGATCTGGCAACTTCTTAAAGAGTGT 1147  
QY 1189 GGGGTTTAAATGATGAACAGAGAACTGGCAACCCATGATCAATCTTACGGAACG 1248  
DB 1148 GGGGTTTAAATGATGAACAGAGAACTGGGCAACCCATGATCAATCTTACGGAACG 1207  
QY 1249 GAAACAGAAAGCCCAAGGCGATGCAAGTGTCTTATGAAGAACCAACCACTGCCAAG 1308  
DB 1208 GAAACAGAAAGCCCAAGGCGATGCAAGTGTCTTATGAAGAACCAACCACTGCCAAG 1267  
QY 1309 GCTGCGTGAATGTTTATGAGAAAGATTTTCAAGGAGCAACTTAAAGTCTCCCTT 1368  
DB 1268 GCTGCGTGAATGTTTATGAGAAAGATTTTCAAGGAGCAACTTAAAGTCTCCCTT 1327







QY	364	GCCTCTATGCAAGCTACGCTGCATATGGCACTCAGCGCTGTTATCCAGCCATATGGGAG	423
Db	10714	GCCTCTATGCAAGCTCAATCTGCATATGGCACTCAGCGCTGTTATCCAGCTTATGGGAG	10773
QY	424	CAGCCAGCAAGCACTGCACTTACAAAGCCGCGAGATGGAAACAAGCCACTGAGACTGT	483
Db	10774	CAGCCAGTAGCACTGCACTTCAAGAGCTGCGAATGGAAACAAGCCCACTGAGACTGT	10833
QY	484	CAACCTCAATCTAGCAACAGGGGGTTACAAACGAGCCAGCTGAGATATGGACAGAGTAC	543
Db	10834	CAACCTCAATCTAGCAACAGGGGGTTACAAACGAGCCAGCTTGGATATGGACAGAGTAC	10893
QY	544	TACAGTTATCCCAAGGTAACCTGGAGCTACCCCATGCAAGCACTGCACTTC	603
Db	10894	TGCAAGTTATCCCAAGGTAACCTGGAGCTACCTCAATGCAAGCACTGCACTTC	10953
QY	604	TACCTCTTACAGCTATTCCTCTACACAGCCGACTATGATCAAGCACTTACT	663
Db	10954	TACCTCTTACAGCTATTCCTCTACAGCCCACTAGTTATGATCAAGCACTTACT	11013
QY	664	CACGAGAACCCATATGGGGAACCGAGACCTATGGACAGCAGTAGTATGGTAAACAA	723
Db	11014	CACGAGAACCCATATGGGGAACCGAGACCTATGGACAGAGTAGTATGGTAAACAA	11073
QY	724	AGCAGTATGGGAGAGGCTCCCACTAGTTACCAACCCCAACTGGATCTTACGCCAA	783
Db	11074	AGCAGTATGGGAGAGGCTCCCACTAGTTACCAACCCCAACTGGATCTTACGCCAA	11133
QY	784	GCTTCGAATCATATATGCCAACAGAGACAGCTATAGGGCACAGAGTTCAATCCGACAG	843
Db	11134	GCTTCGAATCATATATAGCTTAAAGAGACAGACACTATAGGGCACAGAGTTCAATCCGACAG	11193
QY	844	GACCAACCCCAAGTAGCAATGGGTTTATGGGCGAGAGTCTGAGAGATTTTCCGACACGGA	903
Db	11194	GACCAACCCCAAGTAGCAATGGGTTTATGGGCGAGAGTCTGAGAGATTTTCCGACACTGGA	11253
QY	904	GAGAACCGAGCAATGAGTGGCCCTGTATTAACCGGAGCGAGGGGAAGAGGGGGAATTTGATCCT	963
Db	11254	GAGAACCGAGCAATGAGTGGCCCTGTATTAACCTGGAGCGAGGGGAAGAGGGGGAATTTGATCCT	11313
QY	964	GGAGGCAATGAGCAAGAGTGGGGGGGAGAGAGACGGGATGGATATGGGCGCGCTGGAGAG	1023
Db	11314	GGAGGCAATGAGCAAGAGTGGGGGGGAGAGAGATGGGATGGATATGGGCGCGCTGGAGAG	11373
QY	1024	CGAGGTGGCTTCAATTAAGCCTGGTGGACCCATGATGAAAGAAACAAGTCTTGATCTTAGGC	1083
Db	11374	CAAGTTGGCTTCAATTAAGCCTGGTGGACCCATGATGAAAGAAACAAGTCTTGATCTTAGGC	11433
QY	1084	CCACTGTAGATCCAGATGAAGACTCTGACACAGTGCATTTATGTACAAAGATTAAAT	1143
Db	11434	CCACTGTAGATCCAGATGAAGACTCTGACACAGTGCATTTATGTACAAAGATTAAAT	11493
QY	1144	GACAGTGTGACTAGATGATCTGGGAGACTTTTAAAGAGATGGGGTTGTTAAGATG	1203
Db	11494	GACAGTGTGACTAGATGATCTGGTGAACCTTTTAAAGAGATGGGGTTGTTAAGATG	11553
QY	1204	AACAAAGAACTGGGCAACCCATGATCCATATCTAGCTGGAACAAGAAACAGGAAAGCC	1263
Db	11554	AACAAAGAACTGAGCAACCCATGATCCACACTTACCTGGACAAAGAAACAGGAAAGCC	11613
QY	1264	AAAGCGATGCCACAGTGTCTTATGAAGACCCACCTGCGCAAGGCTGCCGTGGAAATGG	1323
Db	11614	AAAGGTATGCCACAGTGTCTGTGAAGACTCACTTACCTGCCAAACCTGCCGTGGAAATGG	11673
QY	1324	TTTGAATGGGAAAGTTTTCAGAGGAGCAACTTAAAGTCCCTTGCTCGGAAGAGCCT	1383
Db	11674	TTTGAATGGGAAAGTTTTCAGAGAGCAACTTAAAGTCTCTTGTCTGGAGAGGCT	11733
QY	1384	CCAATGAACAGTATGCCGGGTGGTCTGCCAACCCTGATAGGGCGAGAGCATGCCACCA	1443
Db	11734	CCAATGAACAGTATGACGGGTGGTATGCCAACCCTATAGGGCGAGAGGATGCCACCA	11793



OY	1444	CTCGGTGAGGATCCAGGAGGCGCCAGGAGGTCCTGAGGGGAAACCAATGGGTGCAATGGGAGGC	1503
Db	11794	CTCTGCGGAGGATCCAGGAGGCGCCAGGAAATCTTGAGGGGAAACCAATGGGTCAATGGGAGGC	11853
OY	1504	CGTGAAGAGATAGAGGAGCTTCCCTCCAAAGAGAACCCGGGGTTCCCGAGGGAACCC	1563
Db	11854	CGTGAAGAGATAGAGGAGGCTTCCCTCCAAAGAGAACCCGAGGTTCCCGAGGGAACAC	11913
OY	1554	TCTGAGGAGGAAACGTCACAGCACCCAGCTGAGACCTGGCAGTGTCCCAATCCGGGTTGT	1623
Db	11914	TCTGAGGAGGAAACGTCACAGCACCAAGCTGAGACAGGCAAGTGTCCCAATCCGGGTTGT	11973
OY	1624	GGAACACGAACTTGCTGGAGAACGAGTGCACCAAGGTAAAGCCCAAGCCGTA	1683
Db	11974	GGAACACGAACTTGCTGGAGAACGAGTGCACCAAGGTAAAGCTCCAAAGCCTGAA	12033
OY	1684	GCGTTCTCTCCGCGCACCTTTCGCGCCCGGGTGTGATCGTGCGAGAGGTGCGCTGTG	1743
Db	12034	GCGTTCTCTCCGCGCACCTTTCGCGCCCGGGTGTGATCATGGCAGAGGTGCGCTGTG	12093
OY	1744	GGCATGCGGGGAGGAAGAGTGGCTCATGATCGTGTGTCCCGGTGAATGTTCA	1803
Db	12094	GGCATGTCGGGAGGAAGAGTGGCTCATGATCATGATGTGTCCCGGTGAATGTTCA	12153
OY	1804	GGTGGCGGTGTGAGACAGAGGTGGCTTCCGTGTGGCCGGGGCAATGACCGAGGTGGC	1863
Db	12154	GGTGGCTGTGTAGACAGAGAGGTGGCTTCTGTGTGTGG-CTGGGCAATGACCGAGGTGGC	12212
OY	1864	TTTGGTGAAGAGAACAGAGTGGCCCTTGAGGGGCGCCCTTGACCTTGTGATGAAACAGATG	1923
Db	12213	TTTGGTGAAGAGAACAGAGGTGGCCCTTGAGGGGCGCCCTTGACCTTGTGATGAAACAGATG	12271
OY	1924	GGAGGAAAGAGAGGAGCCTGAGAGACCTTGAAAAATGGAATAAAGCGACGACCGTCA	1983
Db	12272	GGAGGAAAGAGAGGAGCCTGAGAGACCTTGAAAAATGGAATAAAGCGACGACCTGTCA	12331
OY	1984	GAGCGAGAGATGGCGCCTACTGATGACAGACCCCGCAGA	205
Db	12332	GAGCGAGAGATGAGCCCTACTGATGATCAAGAACCCCGCAGA	12373

ORIGIN

```

URS
source
abbreviations are used to associate primary accession numbers given
in the feature table with their source databases: Em:, EMBL, Sw:,
SWISSPROT, Tr:, TrEMBL, Wp:, WormPEP, information on the WormPEP
database can be found at
http://www.sanger.ac.uk/Projecte/C_elegans/wormep/ This sequence
was generated from part of bacterial clone contigs of human
chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping
Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chrl
RP11-9L18 is from the library RPCT-11.1 constructed by the group of
Pleier de Jong. For further details see
http://www.chori.org/bacpac/home.htm
VECTOR: pBACe3.6
This sequence is the entire insert of clone RP11-9L18 The true left
end of clone RP11-479J7 is at 136234 in this sequence. The true
right end of clone RP11-375H19 is at 73742 in this sequence.

Location/Qualifiers
1..180718
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="1"
/clone="RP11-9L18"
/clone_1lb="RPCT-11.1"
23201..23229
/note="Sequence confirmed by AC011221 and AC018774
sequenced by MIBR."
complement(23230..23297)
/note="[ ] Unsure [x] Misc_feature Then select the text
for the note(s) - [ ] Tandem Repeat [ ] Single clone
region [ ] Forced join [x] Other Add a comment here -
Sequence from AC011221 and AC018774 sequenced by MIBR."
complement(4637..46442)
/note="Single clone region. Sequence from reads from a
short insert library derived from a single pUC clone.
Restriction digest data confirm the assembly."
complement(102116..102124)
/note="767 bases of ISI transposon (X52534) removed here.
This sequence represents the duplicated flanking sequence
of the ISI."
misc_feature
misc_feature
misc_feature

```

RESULT	15
AL596087/c	
LOCUS	
DEFINITION	AL596087 180718 bp DNA linear PRI 10-MAR-2002
ACCESSION	Human DNA sequence from clone RP11-9L18 on chromosome 1, complete sequence.
VERSION	AL596087
KEYWORDS	AL596087.11 GI:19351911
SOURCE	HTG.
ORGANISM	Homo sapiens (human)
REFERENCE	Homo sapiens
AUTHORS	Eukaryotes; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 180718)
TITLE	Almeida,J.
JOURNAL	Direct Submission
COMMENT	Submitted (10-MAR-2002) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: hanguey@sanger.ac.uk Clone requests: clonerequests@sanger.ac.uk On Mar 11, 2002 this sequence version replaced gi:15944185. During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above. This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following

Query Match	92.8%	Score 1879.6	DB 9	Length 180718
Best Local Similarity	96.2%	Pred. No. 0		
Matches 1946	Conservative	0	Mismatches 74	Indels 2
			Gaps	2
QY	4	GGACGTTAGAGAA	CGAGAGGAAAGAGAAATGCGCTCCACGGATTACAGTACTTAT	63
Db	157625	GGACGTTAGAGAA	CGAGAGGAAAGAGAAATATGGATCCATGGATTACAGTACTTAT	157566
QY	64	AGCCAACTGCAGCGCAGACGAGGCTTACAGTCTTACACCGCCAGTCCACTCAAGATAT	123	
Db	157565	AGCCAACTGCAGCGCAGACGAGGCTTACAGTCTTACACACCGCCACTCAAGATAT	157506	
QY	124	GCAAGACACCCACGAGCATATGGGCAACAAAGCTATGGAACCTATGACAGCCACTGAT	183	
Db	157505	GCAAGACACCCACGAGCATATGGGCAACAAAGCTATGGAACCTATGGAACAGCCACTGAT	157446	
QY	184	GTCAGCTTATCCACAGGCTCAGACCATGCGAACCTATGGCAGACCGCTTATGCACTTCT	243	
Db	157445	GTCAGCTTATCCACAGGCTCAGACCATGCAACTTATGGCAGACCGCTTATGCACTTCT	157386	
QY	244	TATGAGACGCTCCCACTGGTTATATCTACTCCACATGCGCCCGACGACATACAGCGCCT	303	
Db	157385	TATGAGACGCTCCCACTGGTTATATCTACTCCACTGCCCGACGACATACAGCGCCT	157328	
QY	304	GTCGAGGGGATATGGCACTGTGTCTTATGATATACCAACACTGTACTAAGTCAACAACCCAG	363	
Db	157325	GTCGAGGGGATATGGCACTGTGTCTTATGATATACCAACACTGTACTAAGTCAACAACCCAG	157266	
QY	364	GGCTTCCTATGCAAGCTCAATCTTGCAATATGGCACTCAGCCGCTTATCAACGCTTATGGGCA	423	
Db	157265	GGCTTCCTATGCAAGCTCAATCTTGCAATATGGCACTCAGCCGCTTATCAACGCTTATGGGCA	157206	

QY	424	CAGCCAGCAGCCACTGCACTTCAACACCGAGAGTGGAAACAAAGCCACTGAGCTAG	483
Db	157205	CAGCCAGTAGCCACTGCACTTCAACACTGAGATGGAAACAGCCCACTGAGCTAG	157146
QY	484	CAACCTCAATCTAGCACAAGGGGTTTCAACACAGCCAGCCTAGATATGGACAGATAC	543
Db	157145	CAACCTCAATCTAGCACAAGGGGTTTCAACACAGCCAGCCTAGATATGGACAGATAC	157086
QY	544	TACAGTTATCCCAGGTAAGTGGAGACTTCCCATGACAGTCACTGACCTCCATCC	603
Db	157085	TGCAAGTTATCCCAGGTAAGTGGAGACTTCCCATGACAGTCACTGACCTCCATCC	157026
QY	604	TACCCCTCTACAGACTATCTCTTCAACAGCCCACTAGTTATGATCAGACATTA	663
Db	157025	TACCCCTCTACAGACTATCTCTTCAACAGCCCACTAGTTATGATCAGACATTA	156966
QY	664	CAGCAGAACACTATGGGCAACGAGACGCTATGACAGCAGATGCTATGGTCAACAA	723
Db	156965	CAGCAGAACACTATGGGCAACGAGACGCTATGACAGCAGATGCTATGGTCAACAA	156906
QY	724	ACACAGTATGGGAGAGGCTCCCACTAGTTACCAACCCCAATGATCTACAGCCAA	783
Db	156905	ACACAGTATGGGAGAGGCTCCCACTAGTTACCAACCCCAATGATCTACAGCCAA	156846
QY	784	GCTCCAAAGTCAATATATGCAACAGACAGACGCTACGGGACAGAGTTCAATCCGACAG	843
Db	156845	GCTCCAAAGTCAATATATGCAACAGACAGACGCTACGGGACAGAGTTCAATCCGACAG	156786
QY	844	GACACCCCAAGTACATGGGGTTTATGGGACAGACTGAGAGATTTTCCCGAACAGAA	903
Db	156785	GACACCCCAAGTACATGGGGTTTATGGGACAGACTGAGAGATTTTCCCGAACAGAA	156726
QY	904	GAGAAACCGAGACATGATGGGCTCTGATTAACCGGGGACAGGGGATTTGATCTG	963
Db	156725	GAGAAACCGAGACATGATGGGCTCTGATTAACCGGGGACAGGGGATTTGATCTG	156666
QY	964	GGAGGCAATAGCAGAGGTGGCGGGGAGAGCAACGGGTGGAAATGGGCAAGCTTGAGAG	1023
Db	156665	GGAGGCAATAGCAGAGGTGGCGGGGAGAGCAACGGGTGGAAATGGGCAAGCTTGAGAG	156606
QY	1024	CGAGGTGGCTTCAATTAAGCTGGTGGACCACTGATGAAAGAACAGATCTTGATCTAGG	1083
Db	156605	CGAGGTGGCTTCAATTAAGCTGGTGGACCACTGATGAAAGAACAGATCTTGATCTAGG	156546
QY	1084	CCACTGTATGATCCAGATAGAAAGCTCTGACAAAGTGCACTTATATGTAACAGGATTAAT	1143
Db	156545	CCACTGTATGATCCAGATAGAAAGCTCTGACAAAGTGCACTTATATGTAACAGGATTAAT	156486
QY	1144	GACAGGTGATCTCTGATGATCTGGAGACTCTTATAGAGGTGGGGTTGTTAAGAT	1203
Db	156485	GACAGGTGATCTCTGATGATCTGGAGACTCTTATAGAGGTGGGGTTGTTAAGAT	156426
QY	1204	AAACAAGAACTGGGCAACCCATGATCAATCTACCTGGACAAAGAAACAGAAAGCCC	1263
Db	156425	AAACAAGAACTGGGCAACCCATGATCAATCTACCTGGACAAAGAAACAGAAAGCCC	156366
QY	1264	AAAGGCGATGCCAGATGCTCTATGAAGACCACTACCTGCAAGCTGCGGTGGAAATGG	1323
Db	156365	AAAGGCGATGCCAGATGCTCTATGAAGACCACTACCTGCAAGCTGCGGTGGAAATGG	156306
QY	1324	TTTGTATGGGAAAGATTTTCAAGGGGACAACTTAAGTCTCCCTGCTCGGAAGAAGCT	1383
Db	156305	TTTGTATGGGAAAGATTTTCAAGGGGACAACTTAAGTCTCTTCTCGGAAGAAGCT	156246
QY	1384	CCAAATGAACATATGCGGGGTGCTTCCACCCCTGAGGGCAGAGGCATGCCAACCA	1443
Db	156245	CCAAATGAACATATGCGGGGTGCTTCCACCCCTGAGGGCAGAGGCATGCCAACCA	156186
QY	1444	CTCCGTGAGAGTCCAGAGAGCCCAAGAGGCTCTGGGGGACCAATGGTGGCAATGGAGGC	1503
Db	156185	CTCCGTGAGAGTCCAGAGAGCCCAAGAGGCTCTGGGGGACCAATGGTGGCAATGGAGGC	156126

QY	1504	CGTGGAGGAGATGAGGAGAGGCTTCCCTCCAAAGAGACCCTGGGTTCCCGAGGAAACCC	1563
Db	156125	CCTGGAGGAGATGAGGAGAGGCTTCCCTCCAAAGAGACCCCAAGGTTCCCGAGGGAACACC	156066
QY	1554	TCTGGAGAGGAAACGTCCAGCACCCGAGCTGGAGACTGGCAGTGTCCCAATCCGGGTTGT	1623
Db	156065	TCTGGAGAGGAAAGTCCAGCACCAAGCTGGAGACAGGCAATGTCTCCAAATCCGGGTTGT	156006
QY	1624	GGAACACCAAGAACTTCGCTCGAGAGACAGAGTCAACCAAGTGTAAAGGCCCAAGCCTGAA	1683
Db	156005	GGAACACCAAGAACTTCGCTCGAGAGACAGAGACCAACAGTGTAAAGGCTCCAAAGCCTGAA	155946
QY	1684	GGCTTCTCTCCGCCCAACCTTTTCGCCCCCGGTTGATCTGTGGCAGAGGTGGCCTTGT	1743
Db	155945	GGCTTCTCTCCGCCCAACCTTTCCCAACCCCGGTTGTGATCATGGCAGAGGTGGCCTTGT	155886
QY	1744	GGCATGCGGGAGGAAAGGTGGCTCACTGATTCGAGTGTATCCCGGTGAAATGTTCAAG	1803
Db	155885	GGCATGTTGGAGAGAAAGAGTGGCTTCATGATTCATGTGTGTCCCGGTGAAATGTTCAAG	155826
QY	1804	GGTGGCCGTGTGTGAGACAGAGGTGGCTTCCGTGTGTGGCCGGGGCATGGACCGAGGTGGC	1863
Db	155825	GGTGGCTGTGTGTGAGACAGAAAGTGGCTTCTGTGTGTGG-CTGGGCAATGGACCGAGTGGC	155767
QY	1864	TTTGTGTGAGGAAAGACAGAGTGGCCCTTGGGGGGCCCCCTTGACCTTTTGTATGGAAACGATG	1923
Db	155766	TTTGTGTGAGGAAAGACAGAGTGGCCCTTGGGGGGCCCC-GGACCTTTTGTATGTAACCAATG	155708
QY	1924	GGAGGAGAGAGAGAGAGACTGTGAGAGACTGTGAAAAAATGATTAAGGCGAGCACTGTAG	1983
Db	155707	GGAGGAGAGAGAGAGAGACTGTGAAAGACTGTGAAAAAACGATTAAGGCGAGCACTGTAG	155648
QY	1984	GAGGCGCAGATCCGGCCCTACTAGATGGAGAGAACCCCGCAGA	2025
Db	155647	GAGGCGAGATCAGCCCTACTAGATGGAGAGAACCCCGCAGA	155606

Search completed: February 20, 2005, 11:50:27  
Job time : 8507.87 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 20, 2005, 00:31:32 ; Search time 982.09 Seconds  
(Without alignments)  
12206.093 Million cell updates/sec

Title: US-10-791-017a-1\_COPY\_8\_2032

Perfect score: 2025  
Sequence: 1 gagcagacgtcagagagacga.....agatgcagagaccgcgcaga 2025

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : N\_Geneseq\_16Dec04:\*  
1: geneseqn19808:\*  
2: geneseqn19908:\*  
3: geneseqn20008:\*  
4: geneseqn2001a:\*  
5: geneseqn2001b:\*  
6: geneseqn2002a:\*  
7: geneseqn2002b:\*  
8: geneseqn2003a:\*  
9: geneseqn2003b:\*  
10: geneseqn2003c:\*  
11: geneseqn2003d:\*  
12: geneseqn2004a:\*  
13: geneseqn2004b:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2025	100.0	2372	13	ADG66032 Human tum
2	2025	100.0	2390	13	ADG66032 Human tum
3	2023.4	99.9	2390	6	ABK84628 Human CDN
4	2023.4	99.9	2390	6	ABK84628 Human CDN
5	2023.4	99.9	2390	6	ABN97274 Gene #377
6	2023.4	99.9	2390	13	ACN40903 Tumour-as
7	2011.4	99.3	2371	2	AAQ50643 Human Ews
8	1986.4	95.9	2388	5	AA662262 CDNA sequ
9	1942.2	92.8	2177	13	ACN37537 Tumour-as
10	1879.6	91.7	2273	6	AA662623 CDNA sequ
11	1856.6	88.3	2188	6	AB199383 Mouse lsc
12	1787.2	81.6	1988	10	ADA53506 Human cod
13	1652.4	80.3	2026	13	ADP66333 Human PRO
14	1625.8	76.9	1807	2	AAQ50646 Human Ews
15	1557.4	49.0	2403	2	AAAT97870 CDNA enco
16	992	40.4	2412	2	AAAT97869 CDNA enco
17	819	39.4	1431	6	AB573253 DNA enco
18	797.6	37.5	856	13	ACN37540 Tumour-as
19	758.4				
20					

## ALIGNMENTS

21	707.4	34.9	845	13	ACN37538	ACN37538 Tumour-as
22	653.6	32.3	1462	4	AA126725	AA126725 Human brc
23	570	28.1	2299	11	ACN88697	ACN88697 Breast ca
24	514.4	25.4	540	4	AA117281	AA117281 Probe #72
25	514.4	25.4	540	4	ABA62106	ABA62106 Human foe
26	514.4	25.4	540	4	AA142050	AA142050 Probe #10
27	514.4	25.4	540	4	ABA29535	ABA29535 Probe #80
28	514.4	25.4	540	4	AAK36327	AAK36327 Human bon
29	514.4	25.4	540	4	AAK10423	AAK10423 Human bra
30	514.4	25.4	540	6	ABS36007	ABS36007 Human gen
31	514.4	25.4	550	9	ACH40478	ACH40478 Human foe
32	508.4	25.1	684	13	ACN37539	ACN37539 Tumour-as
33	488.4	23.7	568	9	ACH22345	ACH22345 Human adu
34	480.2	20.9	454	3	AAAC01659	AAAC01659 Human sec
35	424	20.7	471	2	AAV86051	AAV86051 EST clone
36	419.2	20.3	505	4	AA117153	AA117153 Probe #70
37	411.6	20.3	505	4	ABA61641	ABA61641 Human foe
38	411.6	20.3	505	4	ABA29300	ABA29300 Probe #10
39	411.6	20.3	505	4	AAK35836	AAK35836 Human bon
40	411.6	20.3	505	4	AAK09942	AAK09942 Human bra
41	411.6	20.3	505	4	ABS35546	ABS35546 Human liv
42	411.6	20.3	505	6	ABS10069	ABS10069 Human gen
43	411.6	20.3	505	9	ACH16656	ACH16656 Human adu
44	411.6	19.7	401	9	ACH16656	ACH16656 Human adu
45	399.4					

RESULT 1  
ADG66032 standard; CDNA; 2372 BP.  
ID ADG66032:  
AC ADG66032:  
XX 07-OCT-2004 (first entry)  
DT 07-OCT-2004 (first entry)  
XX Human tumour-associated antigenic target (TAT) cDNA sequence #2904.  
DE human; tumour-associated antigenic target; TAT; cytostatic; gene therapy;  
XX cancer; cell proliferative disorder; gene; ss.  
OS Homo sapiens.  
XX  
XX WO2004060270-A2.  
XX  
XX 22-JUL-2004.  
XX  
XX 15-OCT-2003; 2003WC-US029126.  
XX  
XX 18-OCT-2002; 2002US-0418988P.  
XX  
XX (GETH ) GENENTECH INC.  
XX (WUTD/) WU T D.  
XX (ZHOU/) ZHOU Y.  
XX Wu TD, Zhou Y;  
XX WPI; 2004-534300/51.  
XX  
XX New nucleic acid molecule and encoded polypeptide, for diagnosing,  
XX preventing or treating cell proliferative disorders such as cancer.  
XX  
XX Claim 1; SEQ ID NO 2904; 5504dp; English.  
XX  
XX The present invention describes an isolated tumour-associated antigenic  
XX target (TAT) nucleic acid comprising: (a) any of 4622 nucleotide  
XX sequences (see SEQ ID NO.1 to 4622); (b) the full-length coding region of  
XX (a); (c) the complement of (a) or (b); (d) a sequence that has 80%  
XX sequence identity to (a) - (c); or (e) a sequence that hybridises to (a) -  
XX (c). Also described: (1) an expression vector comprising the above  
XX nucleic acid; (2) a host cell comprising the above expression vector; (3)

a process for producing a polypeptide; (4) an isolated polypeptide comprising: (a) an amino acid sequence encoded by any of the above nucleotide sequences; (b) an amino acid sequence encoded by the full-length coding region of the above nucleotide sequences; or (c) a sequence having at least 80% identical to (a) or (b); (5) a chimeric polypeptide comprising the above polypeptide fused to a heterologous polypeptide; (6) an isolated antibody that binds to the above polypeptide; (7) a process for producing the antibody; (8) an isolated oligopeptide that binds to the above polypeptide; (9) a tumour-associated antigenic target (TAT) binding organic molecule that binds to the above polypeptide; (10) a composition of matter comprising the above (chimeric) polypeptide, antibody, oligopeptide or TAT binding organic molecule, in combination with a carrier; (11) an article of manufacture comprising a container and the composition of matter contained within the container; (12) methods of inhibiting the growth of a cell that expresses the above protein, where the growth of the cell is at least in part dependent upon a growth potentiating effect of the above protein; (13) a method of potentiating effect of the above protein; (14) a method of pharmaceutically treating a mammal having a cancerous tumour comprising cells that expresses the above protein; (15) a method of determining the presence of a protein in a sample suspected of containing the protein described above; (16) methods of diagnosing the presence of a tumour in a mammal; (16) a method for treating or preventing a cell proliferative disorder associated with increased expression or activity of the above protein; and (17) a method of binding an antibody, oligopeptide or organic molecule to a cell that expresses the protein described above. The TAT sequences have cytostatic activities, and can be used in gene therapy. The composition and methods are useful for diagnosing, preventing or treating cancer. The composition is also used for preparing a medicament for the therapeutic treatment or diagnostic detection of a cell proliferative disorder or cancer. The present sequence represents a human TAT cDNA sequence from the present invention.

SQ Sequence 2372 BP; 628 A; 589 C; 668 G; 487 T; 0 U; 0 Other;

Query Match Similarity	100.0%;	Score 2025;	DB 13;	Length 2372;
Best Local Similarity	100.0%;	Pred. 0;		
Matches 2025; Conservative	0;	Mismatches	0;	Gaps 0;

[illegible]

QY	541	AAC	TAC	GAG	TTA	TCC	CC	AG	TAC	CTG	GGG	AG	CT	CA	CC	CA	TGG	AG	CA	GT	CA	CT	CA	CT	CA	600	
Db	548	AAC	TAC	GAG	TTA	TCC	CC	AG	TAC	CTG	GGG	AG	CT	CA	CC	CA	TGG	AG	CA	GT	CA	CT	CA	CT	CA	607	
QY	601	TCT	TAC	CC	CT	CT	CA	TAC	CT	TA	CT	CT	TA	CA	CA	CC	CA	CT	TA	GT	TA	GT	TA	GT	TA	660	
Db	608	TCT	TAC	CC	CT	CT	CA	TAC	CT	TA	CT	CT	TA	CA	CA	CC	CA	CT	TA	GT	TA	GT	TA	GT	TA	667	
QY	661	TC	T	C	A	G	A	G	A	A	C	T	T	A	T	G	G	G	A	A	C	G	A	G	A	G	720
Db	668	TC	T	C	A	G	A	G	A	A	C	T	T	A	T	G	G	G	A	A	C	G	A	G	A	G	727
QY	721	C	A	A	G	C	A	G	T	A	T	G	G	G	C	A	G	A	G	C	T	T	A	T	G	G	780
Db	728	C	A	A	G	C	A	G	T	A	T	G	G	G	C	A	G	A	G	C	T	T	A	T	G	G	787
QY	781	C	A	G	C	T	C	A	T	A	T	T	A	G	C	A	A	G	A	G	C	A	G	C	A	T	840
Db	788	C	A	G	C	T	C	A	T	A	T	T	A	G	C	A	A	G	A	G	C	A	G	C	A	T	847
QY	841	C	A	G	A	C	C	A	C	C	A	T	G	G	A	G	C	A	T	G	G	A	G	C	A	T	900
Db	848	C	A	G	A	C	C	A	C	C	A	T	G	G	A	G	C	A	T	G	G	A	G	C	A	T	907
QY	901	G	A	G	A	G	A	C	C	G	A	T	A	G	T	G	C	C	T	G	A	T	A	C	C	G	960
Db	908	G	A	G	A	G	A	C	C	G	A	T	A	G	T	G	C	C	T	G	A	T	A	C	C	G	967
QY	961	C	G	T	G	A	G	C	A	T	G	A	G	C	T	G	G	G	A	G	A	G	A	G	C	T	1020
Db	968	C	G	T	G	A	G	C	A	T	G	A	G	C	T	G	G	G	A	G	A	G	A	G	C	T	1027
QY	1021	G	A	G	C	A	G	T	G	C	T	T	C	A	A	T	A	A	G	C	T	G	T	G	A	C	1080
Db	1028	G	A	G	C	A	G	T	G	C	T	T	C	A	A	T	A	A	G	C	T	G	T	G	A	C	1087
QY	1081	G	G	C	C	A	C	T	T	A	T	T	C	A	G	A	T	T	C	T	G	A	C	A	A	T	1140
Db	1088	G	G	C	C	A	C	T	T	A	T	T	C	A	G	A	T	T	C	T	G	A	C	A	A	T	1147
QY	1141	A	A	T	G	A	C	A	G	T	G	A	C	T	T	A	T	A	G	C	A	G	T	G	G	T	1200
Db	1148	A	A	T	G	A	C	A	G	T	G	A	C	T	T	A	T	A	G	C	A	G	T	G	G	T	1207
QY	1201	A	T	G	A	C	A	G	A	C	T	G	G	C	A	C	C	A	T	C	A	T	C	T	G	A	1266
Db	1208	A	T	G	A	C	A	G	A	C	T	G	G	C	A	C	C	A	T	C	A	T	C	T	G	A	1267
QY	1261	C	C	C	A	A	G	C	C	A	T	G	C	A	G	C	C	A	C	C	A	C	C	A	C	C	1320
Db	1268	C	C	C	A	A	G	C	C	A	T	G	C	A	G	C	C	A	C	C	A	C	C	A	C	C	1327
QY	1321	T	G	T	T	T	G																				

QY 1621 TGTGAAACCAAGACTTGGCTGGAGAAACAGAGTGCACAGGTGTAAGCCCAAGCCT 1680  
DB 1628 TGTGAAACCAAGACTTGGCTGGAGAAACAGAGTGCACAGGTGTAAGCCCAAGCCT 1687  
QY 1681 GAAGGCTTCCCTCCGCAACCTTTCCGCCCCGGGTGGTGTCTGTGCAAGAGTGCCCT 1740  
DB 1688 GAAGGCTTCCCTCCGCAACCTTTCCGCCCCGGGTGGTGTCTGTGCAAGAGTGCCCT 1747  
QY 1741 GGTGGCATGCGGGAGAAAGAGTGGCTCATGATGTGTGTCTCCGGTGAATGTTC 1800  
DB 1748 GGTGGCATGCGGGAGAAAGAGTGGCTCATGATGTGTGTCTCCGGTGAATGTTC 1807  
QY 1801 AGAGGTGCGGTGGTGAAGACAGAGTGGCTTCGTGTGTGGCGGGCATGGAACGAGT 1860  
DB 1808 AGAGGTGCGGTGGTGAAGACAGAGTGGCTTCGTGTGTGGCGGGCATGGAACGAGT 1867  
QY 1861 GGTGGTGGTGAAGAAACAGAGTGGCTTCGTGTGTGGCGGGCATGGAACGAGT 1920  
DB 1868 GGTGGTGGTGAAGAAACAGAGTGGCTTCGTGTGTGGCGGGCATGGAACGAGT 1927  
QY 1921 ATGGAGAGAAAGAGAGAGAGTGGTGGAGACCTGGAAATGATTAAGCGGACCGT 1980  
DB 1928 ATGGAGAGAAAGAGAGAGAGTGGTGGAGACCTGGAAATGATTAAGCGGACCGT 1987  
QY 1981 CAGGAGGCGAGAGTGGCTTACTAGATGAGAGACCCGACAGA 2025  
DB 1988 CAGGAGGCGAGAGTGGCTTACTAGATGAGAGACCCGACAGA 2032

## RESULT 2

ADSI6277  
ID ADSI6277 standard; cDNA; 2390 BP.

AC ADSI6277;

DT 18-NOV-2004 (first entry)

XX Human cDNA encoding the Ewing sarcoma protein Segid 1.

XX human; 88; gene; Ewing sarcoma; EWS; prostatic cancer; alopecia; acne;

XX hypogonadism; androgen-resistance syndrome; testicular feminisation.

XX Homo sapiens.

XX Key Location/Qualifiers

FT CDS 44..2014

FT /tag= a

FT /product= "Ewing sarcoma protein"

PN BP145190-A1.

PD 08-SEP-2004.

PF 16-FEB-2004; 2004BP-00003422.

PR 04-MAR-2003; 2003DE-01009280.

PR 25-APR-2003; 2003US-0465692P.

XX (SCHD ) SCHERING AG.

XX Obendorf M, Wolf S;

XX WPI, 2004-627861/61.

DR P-PSDB; ADSI6278.

XX Determining the hormonal effects of substances, used to identify

PT pharmaceuticals, e.g. for treatment of androgen receptor dysfunction,

PT from modulating interaction between nuclear receptors and Ewing sarcoma

PT protein.

XX Claim 7; SEQ ID NO 1; 30bp; German.

XX This invention relates to a novel modulators that alter the interaction

CC between the Ewing sarcoma protein (EWS) and its nuclear receptor, as well  
CC as the screening method thereof. Specifically, it refers to determining  
CC and identifying a hormonal effect brought about by test compounds that  
CC modulate either the binding of EWS to the nuclear receptor or the ligand-  
CC induced activity of this receptor. The present invention describes the  
CC nuclear receptors as including oestrogen, progesterone, thyroid hormone,  
CC vitamin D, and retinoic acid receptors, most preferably they are androgen  
CC receptors. Accordingly, these modulators may be used in the development  
CC of pharmaceutical compositions that can diagnose and be used to treat  
CC diseases associated with receptor dysfunction such as prostatic cancer,  
CC alopecia, acne, hypogonadism and androgen-resistance syndrome e.g.  
CC testicular feminisation. This method provides reliable, sensitive,  
CC simple, inexpensive and rapid assessment of the hormonal effects of these  
CC test compounds. This polynucleotide sequence is the cDNA encoding the  
CC human Ewing sarcoma protein of the invention.

XX Sequence 2390 BP; 646 A; 589 C; 668 G; 487 T; 0 U; 0 Other;

Query Match 100.0%; Score 2025; DB 13; Length 2390;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 2025; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAGGACGTTGAGAGAAACAGAGAGAAAGAGAAATGCGTCCACGATTAACAGTACC 60  
DB 8 GAGGACGTTGAGAGAAACAGAGAGAAAGAGAAATGCGTCCACGATTAACAGTACC 67  
QY 61 TATAGCCAGCTGAGGCGAGAGGCTTACAGTCTTACACCGCCAGCCACTCAAGGA 120  
DB 68 TATAGCCAGCTGAGGCGAGAGGCTTACAGTCTTACACCGCCAGCCACTCAAGGA 127  
QY 121 TATGCAACAGCAACCCAGGATATGAGCAACAAAGCTATGGAACCTATGAGACGCCACT 180  
DB 128 TATGCAACAGCAACCCAGGATATGAGCAACAAAGCTATGGAACCTATGAGACGCCACT 187  
QY 181 GATGTCAGCTATACCAAGGCTCAGACCACTGCAACCTATGAGAGACCGCTATGCAACT 240  
DB 188 GATGTCAGCTATACCAAGGCTCAGACCACTGCAACCTATGAGAGACCGCTATGCAACT 247  
QY 241 TCTTATGAGACAGCTCCCACTGGTTATATCTTCCAACTGCCCCCAGGATACAGCAG 300  
DB 248 TCTTATGAGACAGCTCCCACTGGTTATATCTTCCAACTGCCCCCAGGATACAGCAG 307  
QY 301 CCTGTCCAGGGGATATGAGCACTGGTCTTATGATACCAACCACTGCTACAGTACACCAACC 360  
DB 308 CCTGTCCAGGGGATATGAGCACTGGTCTTATGATACCAACCACTGCTACAGTACACCAACC 367  
QY 361 CAGGCTCTTATGAGCTGAGTCTGCACTATGAGCACTCAGCTCTTATCCAGCTATGAGG 420  
DB 368 CAGGCTCTTATGAGCTGAGTCTGCACTATGAGCACTCAGCTCTTATCCAGCTATGAGG 427  
QY 421 CAGCAGCAGCAGCAGCTGCACTGCAACAGCCGAGATGAAACAGCCCACTGAGACT 480  
DB 428 CAGCAGCAGCAGCAGCTGCACTGCAACAGCCGAGATGAAACAGCCCACTGAGACT 487  
QY 481 AGTCAACCTCAATCTGACACAGGGGTTACAAACAGCCCACTGAGATTTGACAGAGT 540  
DB 488 AGTCAACCTCAATCTGACACAGGGGTTACAAACAGCCCACTGAGATTTGACAGAGT 547  
QY 541 AACTCAGATTATCCCAAGGTAACCTGGAGAGTACCCCATGAGCAGCTGCACTGCAACCTCCA 600  
DB 548 AACTCAGATTATCCCAAGGTAACCTGGAGAGTACCCCATGAGCAGCTGCACTGCAACCTCCA 607  
QY 601 TCCTACCTTCTTACAGCTATTCCTTTCACACAGCCGATATGATATGACAGAGCTTAC 660  
DB 608 TCCTACCTTCTTACAGCTATTCCTTTCACACAGCCGATATGATATGACAGAGCTTAC 667  
QY 661 TCTCAGAGAAACCTATGAGGACCAAGAGAGTATGAGACAGAGTACTATGTCACAA 720  
DB 668 TCTCAGAGAAACCTATGAGGACCAAGAGAGTATGAGACAGAGTACTATGTCACAA 727  
QY 721 CAAAGCAGCTATGAGGAGAGAGCTCCCACTAGTTACCAACCCCAACTGATCTTACAGC 780  
DB 728 CAAAGCAGCTATGAGGAGAGAGCTCCCACTAGTTACCAACCCCAACTGATCTTACAGC 787

QY 781 CAAGCTTCAGATCATTTATGACCAACAGACGACGCTTACCGGACAGCAAGTTCAATTCGA 840  
 DB 788 CAAGCTTCAGATCATTTATGACCAACAGACGACGCTTACCGGACAGCAAGTTCAATTCGA 847  
 QY 841 CAGGACCAACCCAGTAGCATGGGTGTTTATGGGACAGAGTCTGAGAGATTTTCCGACCA 900  
 DB 848 CAGGACCAACCCAGTAGCATGGGTGTTTATGGGACAGAGTCTGAGAGATTTTCCGACCA 907  
 QY 901 GAGAGAACCCGAGCATGATGGGCTCTGATTAACCGGGACAGGGAGAGGGAGATTTGAT 960  
 DB 908 GAGAGAACCCGAGCATGATGGGCTCTGATTAACCGGGACAGGGAGAGGGAGATTTGAT 967  
 QY 961 CTGAGAGCATGACAGAGGTGGGCGGGAGAGAGACCGGTGGAATGCGGACGCTGGA 1020  
 DB 968 CTGAGAGCATGACAGAGGTGGGCGGGAGAGAGACCGGTGGAATGCGGACGCTGGA 1027  
 QY 1021 GAGGAGGTGGCTTCAATTAAGCTGTGGAACCCATGATGAAGGACAGATCTTGATCTA 1080  
 DB 1028 GAGGAGGTGGCTTCAATTAAGCTGTGGAACCCATGATGAAGGACAGATCTTGATCTA 1087  
 QY 1081 GGCCCACTGTAGATCCAGATGAAGACTCTGACAAAGTGCATTTATGTAACAAGATTA 1140  
 DB 1088 GGCCCACTGTAGATCCAGATGAAGACTCTGACAAAGTGCATTTATGTAACAAGATTA 1147  
 QY 1141 AATGACAGTGTGATCTTATGATGATCTGGGACGACTTTCTTTAAGAGTGGGTGTTTAAG 1200  
 DB 1148 AATGACAGTGTGATCTTATGATGATCTGGGACGACTTTCTTTAAGAGTGGGTGTTTAAG 1207  
 QY 1201 ATGAAACAAGAACTGGGCAACCCATATCACTACTGACACAGAAACAAGAAAG 1260  
 DB 1208 ATGAAACAAGAACTGGGCAACCCATATCACTACTGACACAGAAACAAGAAAG 1267  
 QY 1261 CCCAAGGCGATGACAGATGTCTTATGAAGACCCACCTGCGAAGGCTGCGTGA 1320  
 DB 1268 CCCAAGGCGATGACAGATGTCTTATGAAGACCCACCTGCGAAGGCTGCGTGA 1327  
 QY 1321 TGGTTTATGAGGAAGATTTTCAAGGAGCAAACTTAAAGTCCCTTGTCCGAAAG 1380  
 DB 1328 TGGTTTATGAGGAAGATTTTCAAGGAGCAAACTTAAAGTCCCTTGTCCGAAAG 1387  
 QY 1381 CCTCAATGAACAGATATGCGGGGTGTCTGCAACCCGATGAGGACAGAGCATGCCCA 1440  
 DB 1388 CCTCAATGAACAGATATGCGGGGTGTCTGCAACCCGATGAGGACAGAGCATGCCCA 1447  
 QY 1441 CCACTCGTGAAGGTCCAGAGGCTCCAGAGGTCTTGGGGAGCCCATGGGTGCGATGGA 1500  
 DB 1448 CCACTCGTGAAGGTCCAGAGGCTCCAGAGGTCTTGGGGAGCCCATGGGTGCGATGGA 1507  
 QY 1501 GCGCGTGAAGATTAAGAGGCTTCCCTCAAGAGAGACCCCGGGGTTCCCGAGGAGAC 1560  
 DB 1508 GCGCGTGAAGATTAAGAGGCTTCCCTCAAGAGAGACCCCGGGGTTCCCGAGGAGAC 1567  
 QY 1561 CCTCTGAGAGAGAAAGTCCAGACCCAGCTGAGATCTGAGTGTCCCAATCCGGGT 1620  
 DB 1568 CCTCTGAGAGAGAAAGTCCAGACCCAGCTGAGATCTGAGTGTCCCAATCCGGGT 1627  
 QY 1621 TGTGAAAACAGAACTTGTCTGAGAAACAGAGTGCACAGTGTAAAGGCCCAAGCCT 1680  
 DB 1628 TGTGAAAACAGAACTTGTCTGAGAAACAGAGTGCACAGTGTAAAGGCCCAAGCCT 1687  
 QY 1681 GAAAGCTTCTCCCGGCAACCTTTCCCGGCGGGGTGTATGCTGTGCAAGAGTGGCCT 1740  
 DB 1688 GAAAGCTTCTCCCGGCAACCTTTCCCGGCGGGGTGTATGCTGTGCAAGAGTGGCCT 1747  
 QY 1741 GGTGCAATGCGGGAGAAAGATGCTTCAATGATCGTGTGTCCCGGTGAAATGTC 1800  
 DB 1748 GGTGCAATGCGGGAGAAAGATGCTTCAATGATCGTGTGTCCCGGTGAAATGTC 1807  
 QY 1801 AGAGGTGGCCGTGTGAGACAGAGGTGCTTCCGTGTGGCCGAGGACATGACCGAGGT 1860  
 DB 1808 AGAGGTGGCCGTGTGAGACAGAGGTGCTTCCGTGTGGCCGAGGACATGACCGAGGT 1867

QY 1861 GGCTTTGATGAGAAAGACAGATGGCCCTGGGGGGCCCCCTTGACCTTTGATGAAACAG 1920  
 DB 1868 GGCTTTGATGAGAAAGACAGATGGCCCTGGGGGGCCCCCTTGACCTTTGATGAAACAG 1927  
 QY 1921 ATGGAGAAAGAAAGAGAGACGTGAGAGACCTTGAAAAATGATTAAGCGACACCGT 1980  
 DB 1928 ATGGAGAAAGAAAGAGAGACGTGAGAGACCTTGAAAAATGATTAAGCGACACCGT 1987  
 QY 1981 CAGGACCGCAGAGATGGCTTACTTATGATGACAGACCCCCGACGA 2025  
 DB 1988 CAGGACCGCAGAGATGGCTTACTTATGATGACAGACCCCCGACGA 2032  
 RESULT 3  
 ABK84628  
 ID ABK84628 standard; cDNA; 2390 BP.  
 AC  
 XX  
 XX  
 DT 14-AUG-2002 (first entry)  
 XX  
 XX  
 DE Human cDNA differentially expressed in granulocytic cells #1199.  
 XX  
 XX Human; ss; granulocytic cell; DNA chip; bacterial infection;  
 KM viral infection; parasitic infection; protozoal infection;  
 KM fungal infection; sterile inflammatory disease; psoriasis;  
 KM rheumatoid arthritis; glomerulonephritis; asthma; thrombosis;  
 KM cardiac reperfusion injury; renal reperfusion injury; ARDS;  
 KM adult respiratory distress syndrome; inflammatory bowel disease;  
 KM Crohn's disease; ulcerative colitis; periodontal disease;  
 KM granulocyte activation; chronic inflammation; allergy.  
 XX  
 OS Homo sapiens.  
 XX  
 XX W0200228999-A2.  
 XX  
 PD 11-APR-2002.  
 XX  
 PF 03-OCT-2001; 2001MO-US030821.  
 XX  
 XX  
 PR 03-OCT-2000; 2000US-0237189P.  
 XX  
 XX  
 PA (GENE-) GENE LOGIC INC.  
 XX  
 PI Beazer-Barclay Y, Weissman SM, Yamaga S, Vockley J;  
 DR WPI; 2002-435328/46.  
 XX  
 XX  
 PT Detecting granulocyte activation by detecting differential expression of  
 PT genes associated with granulocyte activation, which serves as diagnostic  
 PT markers that is useful for monitoring disease states and drug toxicity.  
 XX  
 XX  
 PS Claim 1, SEQ ID NO 1199; 114bp; English.  
 XX  
 XX The invention relates to detecting (M1) granulocyte (GC) activation  
 CC (GCA), by detecting the level of expression of gene(s) (Gs) identified by  
 CC DNA chip analysis as given in the specification, and comparing the  
 CC expression level to an expression level in an unactivated GC, where  
 CC differential expression of Gs is indicative of GCA. Also included are  
 CC modulating (M2) Gs by contacting GC with an agent that alters the  
 CC expression of at least one gene in Gs; (2) screening (M3) for an agent  
 CC capable of modulating GCA or an inflammation (especially chronic) in a  
 CC tissue, an allergic response in a subject, exposure of a subject to a  
 CC pathogen or sterile inflammatory disease using the gene expression  
 CC profile; (3) detecting (M4) an inflammation (especially chronic) in a  
 CC tissue, an allergic response in a subject, exposure of a subject to a  
 CC pathogen or sterile inflammatory disease, by detecting the level of  
 CC expression in a sample of the tissue of gene(s) from Gs, where the level  
 CC of expression of the gene is indicative of inflammation; (4) treating  
 CC (M5) an inflammation (especially chronic) or in a tissue, an allergic  
 CC response in a subject, exposure of a subject to a pathogen or sterile  
 CC inflammatory disease, by contacting a tissue having inflammation with an  
 CC agent that modulates the expression of gene(s) from Gs in the tissue. M1





OY 1661 GGCTTTGGTGAAGAAAGACAGAGTGGCCCTGTGGGGGGCCCTGTGACCTTTGATGAAACAG 1920  
 Db 1668 GGCCTTTGGTGAAGAAAGACAGAGTGGCCCTGTGGGGGGCCCTGTGACCTTTGATGAAACAG 1927  
 OY 1921 ATGGAGAGAAAGAGAGAGACGTGTGAGAACCTTGAAAAATGATTAAGGCGACGACCGT 1980  
 Db 1928 ATGGAGAGAAAGAGAGAGACGTGTGAGAGACCTTGAAAAATGATTAAGGCGACGACCGT 1987  
 OY 1981 CAGGAGCGCAGAGATCGGCCCTTACTAGATTGCAGAGACCCCGCCAGA 2025  
 Db 1988 CAGGAGCGCAGAGATCGGCCCTTACTAGATTGCAGAGACCCCGCCAGA 2032  
 RESULT 4  
 ID ABRK64822 standard, DNA; 2390 BP.  
 XX ABRK64822;  
 AC  
 XX  
 DT 18-JUN-2002 (first entry)  
 XX  
 DE Human benign prostatic hyperplasia gene #717.  
 KW Human; benign prostatic hyperplasia; BPH; prostate cancer; gene; da.  
 RV  
 XX Homo sapiens.  
 OS  
 XX WO200212440-A2.  
 PN 14-FEB-2002.  
 PD  
 PF 07-AUG-2001; 2001WO-US024708.  
 XX  
 PR 07-AUG-2000; 2000US-0223323P.  
 XX 05-JUN-2001; 2001US-00873319.  
 PA (GENE-) GENE LOGIC INC.  
 PA (NITSB) JAPAN TOBACCO INC.  
 PI Munger WE, Kulkarni P, Getzenberg RH, Waga I, Yamamoto J;  
 XX WPI; 2002-257476/30.  
 DR  
 XX  
 PT Identifying drugs for and diagnosing benign prostatic hyperplasia, by  
 PT detecting expression levels of one or more genes in prostate cells from  
 PT patient that are differentially regulated compared to normal prostate  
 PT cells.  
 PS  
 XX Disclosure; Page 399-400; 444pp; English.  
 CC The invention relates to a method of diagnosing (I) the onset or  
 CC progression of benign prostatic hyperplasia (BPH), or screening (II) for  
 CC or identifying an agent that modulates the onset or progression of BPH.  
 CC The method is based on changes in gene expression in BPH tissue isolated  
 CC from patients exhibiting different clinical states of prostate  
 CC hyperplasia as compared to normal prostate tissue. (I) comprises  
 CC detecting the expression levels of one or more genes in prostate cells  
 CC from the subject that are differentially regulated compared to normal  
 CC prostate cells. (II) comprises preparing a first gene expression profile  
 CC of BPH cells or BPH-like cell population, exposing the cells to the  
 CC agent, preparing a second gene expression profile of the agent exposed  
 CC cells, and comparing the first and second gene expression profiles. (I)  
 CC is useful for diagnosing the onset or progression of BPH. (II) is useful  
 CC for identifying an agent that modulates the onset or progression of BPH.  
 CC The methods are useful to present information identifying the expression  
 CC level in a tissue or cells, by comparing the expression level of genes  
 CC given in the specification in the database, and displaying the expression levels  
 CC of expression of gene in the database, and displaying the expression levels  
 CC of at least one gene in the tissue or cell sample compared to the  
 CC expression level in BPH. Agents using (II) are useful for treating BPH or  
 CC prostate cancer. ABRK4106-ABRK6860 represent human benign prostatic  
 CC hyperplasia gene sequences of the invention

XX	Sequence	2390	BE; 645	A; 589	C; 668	G; 488	T; 0	U; 0	Other;	
Query Match	99.9%;	Score 2023.4;	DB 6;	Length 2390;						
Best Local Similarity	100.0%;	Pred. No. 0;								
Matches 2024;	Conservative	0;	Mismatches	1;	Indels	0;	Gaps	0;		
QY	1	GACGGA	CGTTGAGAGAA	CAAGAGAGAA	GGAGAGAAAA	TGGCGTCC	ACGGATTTA	CAGTACC	60	
DB	8	GACGGA	CGTTGAGAGAA	CAAGAGAGAA	GGAGAGAAAA	TGGCGTCC	ACGGATTTA	CAGTACC	67	
QY	61	TATAGC	CAAGCTGACGGC	AGCAGAGGGCT	ACAGTGTTC	ACCGCCCA	GCCCACTCA	AGGA	120	
DB	68	TATAGC	CAAGCTGACGGC	AGCAGAGGGCT	ACAGTGTTC	ACCGCCCA	GCCCACTCA	AGGA	127	
QY	121	TATGCA	CAGACCA	CCGAGCAT	TGGGCA	CAAA	AGCTAT	TGGAAC	CTAT	180
DB	128	TATGCA	CAGACCA	CCGAGCAT	TGGGCA	CAAA	AGCTAT	TGGAAC	CTAT	187
QY	181	GATGTC	AGCTAT	TCCAGGCT	CAAGACCA	CTGCAACCT	ATGGGAGAC	CCGCTAT	ATGCACT	240
DB	188	GATGTC	AGCTAT	TCCAGGCT	CAAGACCA	CTGCAACCT	ATGGGAGAC	CCGCTAT	ATGCACT	247
QY	241	TCTTAT	GAGACAG	CGCTCC	CACTGTGT	TATATCT	CCAA	CTGCCCC	CCAGAT	300
DB	248	TCTTAT	GAGACAG	CGCTCC	CACTGTGT	TATATCT	CCAA	CTGCCCC	CCAGAT	307
QY	301	CTGTG	CAAGGGGT	TGGCACT	GTGTGT	TATAT	ATACCA	CACTGT	ATAC	360
DB	308	CTGTG	CAAGGGGT	TGGCACT	GTGTGT	TATAT	ATACCA	CACTGT	ATAC	367
QY	361	CAGGCG	CTCTAT	GAGCTCA	GTCTGAT	TATGGG	CACTCAG	CGCTGT	TATCC	420
DB	368	CAGGCG	CTCTAT	GAGCTCA	GTCTGAT	TATGGG	CACTCAG	CGCTGT	TATCC	427
QY	421	CAGGAC	CAGACG	CACTGCA	CTTACA	CAAGA	CCGAGANT	GAACA	CAAGCC	480
DB	428	CAGGAC	CAGACG	CACTGCA	CTTACA	CAAGA	CCGAGANT	GAACA	CAAGCC	487
QY	481	AGTCA	ACCTCA	TATAGCA	CAGGGGGT	TACA	CAAGCC	CAGCCT	ATAT	540
DB	488	AGTCA	ACCTCA	TATAGCA	CAGGGGGT	TACA	CAAGCC	CAGCCT	ATAT	547
QY	541	AACTCA	CACTTAT	CCCCAG	TACCTGGG	AGCTAC	CCCCAT	GGACAG	CACTG	600
DB	548	AACTCA	CACTTAT	CCCCAG	TACCTGGG	AGCTAC	CCCCAT	GGACAG	CACTG	607
QY	601	TCTTAC	CCCTCT	ACAGCT	ATTTCT	CTTACA	CAGCCGA	CTAGTT	ATGAT	660
DB	608	TCTTAC	CCCTCT	ACAGCT	ATTTCT	CTTACA	CAGCCGA	CTAGTT	ATGAT	667
QY	661	TCTGAC	GAGAAC	CACCTAT	TGGGCA	CCGAG	CAGCTAT	GGACAG	CAGAGT	720
DB	668	TCTGAC	GAGAAC	CACCTAT	TGGGCA	CCGAG	CAGCTAT	GGACAG	CAGAGT	727
QY	721	CAAA	GACGCT	ATGGG	CAGACG	CTTCCCA	CTAT	AGTTAC	CA	780
DB	728	CAAA	GACGCT	ATGGG	CAGACG	CTTCCCA	CTAT	AGTTAC	CA	787
QY	781	CAAG	CTCCAA	GTCAAT	TATAGC	CAACAG	AGCAGCT	ACGGG	CAGAGT	840
DB	788	CAAG	CTCCAA	GTCAAT	TATAGC	CAACAG	AGCAGCT	ACGGG	CAGAGT	847
QY	841	CAGAG	CAACCC	CAATAG	ATGGG	ATTTTA	TGGG	CAGAGT	CTTGG	900
DB	848	CAGAG	CAACCC	CAATAG	ATGGG	ATTTTA	TGGG	CAGAGT	CTTGG	907
QY	901	GAGAG	GAAC	CCGAG	AGAT	ATG	TGG	CCCTG	AT	960
DB	908	GAGAG	GAAC	CCGAG	AGAT	ATG	TGG	CCCTG	AT	967
QY	961	CGTGA	AGGCAT	GACAG	AGTGGG	CGGGAG	AGAGAG	CGCG	TGGA	1020

Db 968 CGTGAAGCATGAGAGAGTGGCGGGGAGAGAGACGGGTGGAATGGCAGGCGCTGA 1027  
Qy 1021 GAGGAGGTGGCTTCAATAGCCTGGTGGACCCATGGATGAAGACCAAGATCTTGATTA 1080  
Db 1028 GAGGAGGTGGCTTCAATAGCCTGGTGGACCCATGGATGAAGACCAAGATCTTGATTA 1087  
Qy 1081 GGCCCACTGTGATTCAGATGAAGACTGTGACAAAGTGAATTTTATGTAAGAGATTA 1140  
Db 1088 GGCCCTCTGTGATTCAGATGAAGACTGTGACAAAGTGAATTTTATGTAAGAGATTA 1147  
Qy 1141 AATGACAGTGTGATCTAGATGATCTGACAGACTTCTTTAAGCAGTGTGGGTGTTAG 1200  
Db 1148 AATGACAGTGTGATCTAGATGATCTGACAGACTTCTTTAAGCAGTGTGGGTGTTAG 1207  
Qy 1201 ATGAAACAAGAACTGGGCAACCAATGATCAATCTACTGTAACAGAAACAGAAAG 1260  
Db 1208 ATGAAACAAGAACTGGGCAACCAATGATCAATCTACTGTAACAGAAACAGAAAG 1267  
Qy 1261 CCCAAAGGCGATGCGACAGTGTCTTATGAAAGACCCACCTGCAAGGCTGGCTGGA 1320  
Db 1268 CCCAAAGGCGATGCGACAGTGTCTTATGAAAGACCCACCTGCAAGGCTGGCTGGA 1327  
Qy 1321 TGGTTTATGAGGAAGATTTTCAAGGAGCAAACTTAAAGTCTCCCTTGTCTGGAAG 1380  
Db 1328 TGGTTTATGAGGAAGATTTTCAAGGAGCAAACTTAAAGTCTCCCTTGTCTGGAAG 1387  
Qy 1381 CCTTCAATGAACAGTATGCGGGGTGTCTGTCACCCCGTGAAGGCAAGGATCCACA 1440  
Db 1388 CCTTCAATGAACAGTATGCGGGGTGTCTGTCACCCCGTGAAGGCAAGGATCCACA 1447  
Qy 1441 CCACTCGTGAAGTCCAGAGAGGCCAGAGAGTCTGGGGGACCAATGGTGGATGGA 1500  
Db 1448 CCACTCGTGAAGTCCAGAGAGGCCAGAGAGTCTGGGGGACCAATGGTGGATGGA 1507  
Qy 1501 GGCCTGTGAGAGATAGAGAGGCTTCTCTTCAAGAGAGACCCCGGGTTCCTCGAGGAG 1560  
Db 1508 GGCCTGTGAGAGATAGAGAGGCTTCTCTTCAAGAGAGACCCCGGGTTCCTCGAGGAG 1567  
Qy 1561 CCTTGTGAGAGAGAAAGTTCAGACACCGAGCTGGAAGATGAGCTGCTCCATCCGGGT 1620  
Db 1568 CCTTGTGAGAGAGAAAGTTCAGACACCGAGCTGGAAGATGAGCTGCTCCATCCGGGT 1627  
Qy 1621 TGTGAAACAGAACTTGCGCTGGAAGACAGAGTGAACCAAGTGAAGGCCCAAGACCT 1680  
Db 1628 TGTGAAACAGAACTTGCGCTGGAAGACAGAGTGAACCAAGTGAAGGCCCAAGACCT 1687  
Qy 1681 GAAAGCTTCTCCCGCACCTTTCGCGCCCGGGTGTGATCTGTGCAAGAGTGGCCCT 1740  
Db 1688 GAAAGCTTCTCCCGCACCTTTCGCGCCCGGGTGTGATCTGTGCAAGAGTGGCCCT 1747  
Qy 1741 GGTGCAATGCGGGAGAGAGAGTGGCTCAATGATCTGTGTGCTCCCGGTGAATGTC 1800  
Db 1748 GGTGCAATGCGGGAGAGAGAGTGGCTCAATGATCTGTGTGCTCCCGGTGAATGTC 1807  
Qy 1801 AGAGGTGGCGGTGTGAGACAGAGGTGGCTTCCGTGTGTGGCCGGGCAATGACCGAGT 1860  
Db 1808 AGAGGTGGCGGTGTGAGACAGAGGTGGCTTCCGTGTGTGGCCGGGCAATGACCGAGT 1867  
Qy 1861 GGTCTTGTGAGAGAGACAGAGTGGCTTGGGGGGGCCCTTGAACCTTTGATGAACAG 1920  
Db 1868 GGTCTTGTGAGAGAGACAGAGTGGCTTGGGGGGGCCCTTGAACCTTTGATGAACAG 1927  
Qy 1921 ATGGAGGAAG 1980  
Db 1928 ATGGAGGAAG 1987  
Qy 1981 CAGGAGGCGAGAGATGGGCGCTTAATAGATGAGAGAGAGAGAGAGAGAGAGAGAG 2032  
Db 1988 CAGGAGGCGAGAGATGGGCGCTTAATAGATGAGAGAGAGAGAGAGAGAGAGAGAG

ID ABN97274 standard; DNA; 2390 BP.  
XX  
AC ABN97274;  
XX  
DT 13-AUG-2002 (first entry)  
XX  
DE Gene #3772 used to diagnose liver cancer.  
XX  
KM Gene; liver cancer; der, hepatocellular carcinoma; hepatotropic;  
KM metastatic liver tumor; cytostatic; expression profile; disease state;  
KM disease progression; drug toxicity; drug efficacy; drug metabolism.  
OS Homo sapiens.  
XX  
PN M0200229103-A2.  
XX  
PD 11-APR-2002.  
XX  
PF 02-OCT-2001; 2001MO-US030589.  
XX  
PR 02-OCT-2000; 2000US-0237054P.  
XX  
PA (GENE-) GENE LOGIC INC.  
XX  
PI Horne D, Alvares C, Peres-da-Silva S, Vockley JG;  
XX  
DR WPI; 2002-426119/45.  
XX  
PT Diagnosis and detecting the progression of liver cancer, hepatocellular  
PT carcinoma or metastatic liver tumor in a patient, involves detecting the  
PT level of expression of two or more genes in a liver tissue sample.  
XX  
PS Claim 1; SEQ ID NO 3772; 298bp; English.  
XX  
CC The invention relates to a novel method for diagnosing and detecting the  
CC progression of liver cancer, hepatocellular carcinoma or metastatic liver  
CC tumor in a patient, and differentiating metastatic liver cancer from  
CC hepatocellular carcinoma in a patient, involving detecting the level of  
CC expression of two or more genes represented in ABN93503-ABN97455 in a  
CC tissue sample. The method of the invention has hepatotropic, and  
CC cytoskeletal activity. The method is useful for diagnosing and detecting  
CC the progression of liver cancer, hepatocellular carcinoma and metastatic  
CC liver carcinoma in a patient. The method is useful for identifying  
CC expression profiles which serve as useful diagnostic markers as well as  
CC markers that can be used to monitor disease states, disease progression,  
CC drug toxicity, drug efficacy and drug metabolism. Note: The sequence data  
CC for this patent did not form part of the printed specification, but was  
CC obtained in electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 2390 BP; 645 A; 589 C; 668 G; 488 T; 0 U; 0 Other;  
Query Match 99.9%; Score 2023.4; DB 6; Length 2390;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 2024; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
Qy 1 GACGAGCTTGTGAGAGAACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 60  
Db 8 GACGAGCTTGTGAGAGAACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 67  
Qy 61 TATAGCCAGCTGACAGGCGACGAGGCTTACAGTGTCTTACACCGGCCACGCTCAAGGA 120  
Db 68 TATAGCCAGCTGACAGGCGACGAGGCTTACAGTGTCTTACACCGGCCACGCTCAAGGA 127  
Qy 121 TATGACAGACGACCCGAGGATATGGGCAACAAAGCTATGGAAGCTATGAGAGGCCACT 180  
Db 128 TATGACAGACGACCCGAGGATATGGGCAACAAAGCTATGGAAGCTATGAGAGGCCACT 187  
Qy 181 GATGTAGCTATACCAAGGCTCAGACCACTCAAGCTATGGGAGAGCCGCTATGCAACT 240  
Db 188 GATGTAGCTATACCAAGGCTCAGACCACTCAAGCTATGGGAGAGCCGCTATGCAACT 247  
Qy 241 TCTTATGAGACAGCTCCCACTGTGTTATATCTCAACTGCCCCCAGGATACAGCAG 300

Db 248 TCTTATGAGCAGCTCCCACTGGTTATCTCACTGCCCCCGAGGATACAGCAG 307  
Qy 301 CTTGTCCAGGGGATATGGCACTGGTCTTATGATACCACTGCTTACATGACCAACC 360  
Db 308 CTTGTCCAGGGGATATGGCACTGGTCTTATGATACCACTGCTTACATGACCAACC 367  
Qy 361 CAGGCTCTTATGAGCTCAGTCTGATATGGCACTGAGCTGCTTATGAGCTTATGG 420  
Db 368 CAGGCTCTTATGAGCTCAGTCTGATATGGCACTGAGCTGCTTATGAGCTTATGG 427  
Qy 421 CAGGAGCAGAGCAGCTGCACTTACAGACCGAGGATGGAAACAGCCCACTGAGACT 480  
Db 428 CAGAGCAGAGCAGCTGCACTTACAGACCGAGGATGGAAACAGCCCACTGAGACT 487  
Qy 481 AGTACACTTCAATTAGACAGAGGGGTTTCAACAGCCCACTGAGATATGACAGAGT 540  
Db 488 AGTCACTCTCAATTAGACAGAGGGGTTTCAACAGCCCACTGAGATATGACAGAGT 547  
Qy 541 AACTACAGTTATCCCAAGTACCTGGAGCTACCCCACTGAGCCAGTCACTGACCTCCA 600  
Db 548 AACTACAGTTATCCCAAGTACCTGGAGCTACCCCACTGAGCCAGTCACTGACCTCCA 607  
Qy 601 TCTTACCTCTCTACAGCTATTCCTTACACAGCCGACTAGTTATGATCAGAGAGTTAC 660  
Db 608 TCTTACCTCTCTACAGCTATTCCTTACACAGCCGACTAGTTATGATCAGAGAGTTAC 667  
Qy 661 TCTTACAGAAACCTTATGGCAACCGAGCCTATGAGCAGAGAGTATGTTGCA 720  
Db 668 TCTTACAGAAACCTTATGGCAACCGAGCCTATGAGCAGAGAGTATGTTGCA 727  
Qy 721 CAAAGCAGCTATGGGAGCAGCTCCCACTAGTTACCAACCCCAATGATCTTACAGC 780  
Db 728 CAAAGCAGCTATGGGAGCAGCTCCCACTAGTTACCAACCCCAATGATCTTACAGC 787  
Qy 781 CAACTCCAACTCAATATAGCAACAGAGCAGCTACGGGAGCAGAGTTCTTACGA 840  
Db 788 CAACTCCAACTCAATATAGCAACAGAGCAGCTACGGGAGCAGAGTTCTTACGA 847  
Qy 841 CAGAGCAACCCCACTACATGGGTTTATGGGAGAGTCTGAGAGATTTCCGAGCA 900  
Db 848 CAGAGCAACCCCACTACATGGGTTTATGGGAGAGTCTGAGAGATTTCCGAGCA 907  
Qy 901 GAGAGAAACCGAGCAGTATGAGTGGCTTGAATACCGGGGAGGAGGAGGAGTTGAT 960  
Db 908 GAGAGAAACCGAGCAGTATGAGTGGCTTGAATACCGGGGAGGAGGAGGAGTTGAT 967  
Qy 961 CGTGAAGCATGAGCAGAGTGGGCGGGAGAGAGACGCGGTGGAATGGGAGGCTGGA 1020  
Db 968 CGTGAAGCATGAGCAGAGTGGGCGGGAGAGAGACGCGGTGGAATGGGAGGCTGGA 1027  
Qy 1021 GAGGAGTGGCTTCAATTAAGCTGGTGGACCCATGATGAAGAACCAAGATCTTGAATCTA 1080  
Db 1028 GAGGAGTGGCTTCAATTAAGCTGGTGGACCCATGATGAAGAACCAAGATCTTGAATCTA 1087  
Qy 1081 GAGCCACCTGATAGTCCAGATGAGAGCTCTGACAAAGTGTATTTATGATCAAGATTA 1140  
Db 1088 GAGCCCTCTGATAGTCCAGATGAGAGCTCTGACAAAGTGTATTTATGATCAAGATTA 1147  
Qy 1141 AATGACAGTGTAGCTTATGATATCTGGCAGACTTTTAAAGCAGTGTGGGTTGTTAG 1200  
Db 1148 AATGACAGTGTAGCTTATGATATCTGGCAGACTTTTAAAGCAGTGTGGGTTGTTAG 1207  
Qy 1201 AATGACAGAGAACTTGGGCAACCCATGATCCACATCTTACGAGCAAGAAACAGAAAG 1260  
Db 1208 AATGACAGAGAACTTGGGCAACCCATGATCCACATCTTACGAGCAAGAAACAGAAAG 1267  
Qy 1261 CCCAAAGGAGTGCAGAGTGTCTTATGAAGACCAACCACTGCAAGGCTGCGGTGGA 1320  
Db 1268 CCCAAAGGAGTGCAGAGTGTCTTATGAAGACCAACCACTGCAAGGCTGCGGTGGA 1327  
Qy 1321 TGGTTATGAGGAAAGATTTTCAAGGAGCAACTTAAAGTCTCTGCTGGAAGAG 1380

Db 1328 TGGTTATGAGGAAAGATTTTCAAGGAGCAACTTAAAGTCTCTGCTGGAAGAG 1387  
Qy 1381 CTTCCAAATGAACGATATGGGGGTGTTCTCCACCCCGTGAAGGAGCAGAGCAGCA 1440  
Db 1388 CTTCCAAATGAACGATATGGGGGTGTTCTCCACCCCGTGAAGGAGCAGAGCAGCA 1447  
Qy 1441 CCACTCCGTGAGAGTCCAGAGGAGCCAGAGAGTCTTGGGAGAACCATGAGTCCAGTGG 1500  
Db 1448 CCACTCCGTGAGAGTCCAGAGGAGCCAGAGAGTCTTGGGAGAACCATGAGTCCAGTGG 1507  
Qy 1501 GGCCTGAGAGAGATAGAGAGGCTTCCCTTCAAGAGAACCCCGGGTTCCTCCAGGAGAC 1560  
Db 1508 GGCCTGAGAGAGATAGAGAGGCTTCCCTTCAAGAGAACCCCGGGTTCCTCCAGGAGAC 1567  
Qy 1561 CCCTCTGAGAGAGAAACGTCAGACCCGAGCTGAGACCTGAGATGTCCTCAATCCGGGT 1620  
Db 1568 CCCTCTGAGAGAGAAACGTCAGACCCGAGCTGAGACCTGAGATGTCCTCAATCCGGGT 1627  
Qy 1621 TGTGAAACCAAGAACTTCCGCTGAGAAACAGAGTGCACCAAGTGAAGGCCCAAGGCT 1680  
Db 1628 TGTGAAACCAAGAACTTCCGCTGAGAAACAGAGTGCACCAAGTGAAGGCCCAAGGCT 1687  
Qy 1681 GAAAGCTTCTCCGCAACCTTTCGCGCCCGGGTGTGATGTTGAGAGAGTGCCT 1740  
Db 1688 GAAAGCTTCTCCGCAACCTTTCGCGCCCGGGTGTGATGTTGAGAGAGTGCCT 1747  
Qy 1741 GGTGCAATGGGGAGAGAAAGGTGCTTCAATGATGATGATGATGATGATGATGATGATG 1800  
Db 1748 GGTGCAATGGGGAGAGAAAGGTGCTTCAATGATGATGATGATGATGATGATGATGATG 1807  
Qy 1801 AGAGTGGCCGTGTTGAGAGACAGAGGTGCTTCCGTGTTGAGCCGAGGATGAGCCAGT 1860  
Db 1808 AGAGTGGCCGTGTTGAGAGACAGAGGTGCTTCCGTGTTGAGCCGAGGATGAGCCAGT 1867  
Qy 1861 GGTGTTGTTGAGAGAAACAGAGTGGCTTGGGGGGCCCTTGAACCTTTGATGAACAG 1920  
Db 1868 GGTGTTGTTGAGAGAAACAGAGTGGCTTGGGGGGCCCTTGAACCTTTGATGAACAG 1927  
Qy 1921 ATGGAGAGAAAGAGAGAGAGTGGAGACCTGAAATTAAGTAAAGCGAGCAGCT 1980  
Db 1928 ATGGAGAGAAAGAGAGAGAGTGGAGACCTGAAATTAAGTAAAGCGAGCAGCT 1987  
Qy 1981 CAGAGGCGCAGAGATCGCCCTTACTAGATGACAGAGACCCCGCAGA 2025  
Db 1988 CAGAGGCGCAGAGATCGCCCTTACTAGATGACAGAGACCCCGCAGA 2032

RESULT 6  
ACN40903  
ID ACN40903 standard; cDNA; 2390 BP.  
XX  
AC ACN40903;  
XX  
DT 18-NOV-2004 (first entry)  
XX  
DE Tumour-associated antigenic target (TAT) cDNA DNA269830, SEQ ID NO:5989.  
XX  
KW Tumour-associated antigenic target; TAT; human; overexpression; cancer;  
KW tumour; diagnosis; cell proliferative disorder; breast cancer;  
KW colorectal cancer; lung cancer; ovarian cancer; liver cancer;  
KW central nervous system cancer; bladder cancer; pancreatic cancer;  
KW cervical cancer; melanoma; leukemia; hybridization probe;  
KW chromosome identification; chromosome mapping; gene mapping;  
KW gene therapy; cytostatic; gene; ss.  
XX  
OS Homo sapiens.  
XX  
PN MO2004030615-A2.  
XX  
PD 15-APR-2004.  
XX  
PF 29-SEP-2003; 2003MO-US028547.  
XX





Oy	1561	CCCTTCGAGGAGGAAAAGTCCACAGACGAACTGGAGATCGCAGTGTCCCAATCCGGGT	1620
Db	1568	CCCTTCGAGGAGGAAAAGTCCACAGACGAACTGGAGATCGCAGTGTCCCAATCCGGGT	1627
Oy	1621	TGTGGAAACCAAACTTCGCTCGAGAAACAAGTGTCAACAAGTGTAAAGGCCCCAAAGCTT	1680
Db	1628	TGTGGAAACCAAACTTCGCTCGAGAAACAAGTGTCAACAAGTGTAAAGGCCCCAAAGCTT	1687
Oy	1681	GAAGGCTTCCTCCGCGCACCCCTTTCGCGCCCGGGTGTGATCGTGGCAGAGTGGCCCT	1740
Db	1688	GAAGGCTTCCTCCGCGCACCCCTTTCGCGCCCGGGTGTGATCGTGGCAGAGTGGCCCT	1747
Oy	1741	GGTGGCATGCGGGAGAGAAAGTGGCTCATAGATCGTGTGTGTCCCGGTGAAATGTTT	1800
Db	1748	GGTGGCATGCGGGAGAGAAAGTGGCTCATAGATCGTGTGTGTCCCGGTGAAATGTTT	1807
Oy	1801	AGAGTGTGCGCTGTGTGGAGACAAGATGTGGCTTCGTGTGTGCGCCGGGGACATGAAACGAGGT	1866
Db	1808	AGAGTGTGCGCTGTGTGGAGACAAGATGTGGCTTCGTGTGTGCGGGGGACATGAAACGAGGT	1867
Oy	1861	GCGTTTGTGTGAGGAAAGACGAGAGTGGCCCTGTGGGGGCCCTCTTGACCTTTGATGGAAAG	1920
Db	1868	GCGTTTGTGTGAGGAAAGACGAGAGTGGCCCTGTGGGGGCCCTCTTGACCTTTGATGGAAAG	1927
Oy	1921	ATGGAGGAAAGAAAGAGAGACGTGTGAGGACTTGGAAAAATAGATAAAGCGAGACCTGT	1980
Db	1928	ATGGAGGAAAGAAAGAGAGACGTGTGAGGACTTGGAAAAATAGATAAAGCGAGACCTGT	1987
Oy	1981	CAGAGCGCAGAGATCGGCCCTTACTAGATGACAGAGACCCCGCAGA	2025
Db	1988	CAGAGCGCAGAGATCGGCCCTTACTAGATGACAGAGACCCCGCAGA	2032

```

RESULT 7
AAQ50643
ID AAQ50643 standard; cDNA; 2371 BP.
XX
AC AAQ50643;
XX
DT 25-MAR-2003 (revised)
DT 26-MAY-1994 (first entry)
DE Human Ews gene clone B1AC5 from foetal brain cDNA library.
XX
XX chromosomal translocation; chimeric; chimaeric; Ewing sarcoma; Ews gene;
KW malignant melanoma; hum-fil-1;
KW primitive peripheral neuroectodermal tumour; human chromosome 11;
KW human chromosome 22; ss.
XX
OS Homo sapiens.
XX
FH Key location/Qualifiers
FT 25..1995
FT /*tag= a
FT /product= "EWS_protein"
FT /transl_except= pos:1729..1731; aa:Val
FT 2143..2148
FT /*tag= b
FT polyA_signal
FT 2331..2336
FT /*tag= c
XX
XX WO9323549-A2.
XX
XX 25-NOV-1993.
XX
XX 19-MAY-1993; 93WO-FR000494.
XX
XX 20-MAY-1992; 92FR-00006123.
XX
XX (CNRS ) CNRS CENT NAT RECH SCI.
XX
XX Aurias A, Delattre O, Desmaza C, Melot T, Peter M, Plougaetel B,
PI

```

PI	Thomas G.	Zucman J;
XX	WPI; 1993-386580/48.	
DR	P-PSDB; AAR44555.	
XX	New nucleic acid of EMS gene and its hybrid(s) - contg. gene sequence	
PT	involved in chromosomal trans-location, also derived mRNA, probes, fusion	
PT	proteins etc., for diagnosis and treatment of Ewing sarcoma and melanoma.	
XX		
PS	Disclosure; Fig 6; 123pp; French.	
XX	The probes 22R3 and 22R12 were used to screen a human foetal brain cDNA	
CC	library (Stratagene cat.# 936206). The clone BFLA5 was identified and	
CC	sequenced. It represents the entire coding region and 3'-UTR of the Ews	
CC	gene. (Updated on 25-MAR-2003 to correct RN field.)	
XX		
SQ	Sequence 2371 BP; 639 A; 587 C; 659 G; 486 T; 0 U; 0 Other;	
	Query Match	99.3%; Score 2011.4; DB 2; Length 2371;
	Best Local Similarity	100.0%; Pred. No. 0;
	Matches 2012; Conservative	0; Mismatches 1; Indels 0; Gaps 0;
OY	13	GAGAACGAGAGAAGAGAAATATGGCGCTCACGGAATTACAGTACTTAATGCCAGCT 72
Db	1	GAGAACGAGAGAAGAGAAATATGGCGCTCACGGAATTACAGTACTTAATGCCAGCT 60
OY	73	GCAGCGCAGCAGGGGCTACAGTCGTCTTACAACCGCCCAAGCCACTCAAGATATGCACAGCC 132
Db	61	GCAGCGCAGCAGGGGCTACAGTCGTCTTACAACCGCCCAAGCCACTCAAGATATGCACAGCC 120
OY	133	ACCAGGCAATATGGGCAACAAAGCTATGGAACCTATGGACAGCCCACTGATGTGACTAT 192
Db	121	ACCAGGCAATATGGGCAACAAAGCTATGGAACCTATGGACAGCCCACTGATGTGACTAT 180
OY	193	ACCAGGCTCAGACCACTGCACACTATGGGAGACCGGCTATGCACTTTCTTAGACAG 252
Db	181	ACCAGGCTCAGACCACTGCACACTATGGGAGACCGGCTATGCACTTTCTTAGACAG 240
OY	253	CCTCCCACTGGTTATATCTACCTCCAATGCCCCCAAGGATACAGCCAGCCCTGTCCAGGG 312
Db	241	CCTCCCACTGGTTATATCTACCTCCAATGCCCCCAAGGATACAGCCAGCCCTGTCCAGGG 300
OY	313	TATGGCACTGGTGCTATATGATACCAACCACTGCTCAAGTACACCAACCCAGGCTCTAT 372
Db	301	TATGGCACTGGTGCTATATGATACCAACCACTGCTCAAGTACACCAACCCAGGCTCTAT 360
OY	373	GCAGTCACTGTGCATATATGCACTCAGGCTGCTTATCCAGGCTATGGGACAGGCAAGCA 432
Db	361	GCAGTCACTGTGCATATATGCACTCAGGCTGCTTATCCAGGCTATGGGACAGGCAAGCA 420
OY	433	GCCACTGCACCTTACAAAGCCGAGATGGAACAAAGCCCATGGAACCTAGTCAACTCAA 492
Db	421	GCCACTGCACCTTACAAAGCCGAGATGGAACAAAGCCCATGGAACCTAGTCAACTCAA 480
OY	493	TCTAGCAACAGGGGGGTTTCAACAGCCGAGCTAGGATATGACAGAGTAACTTACAGTTAT 552
Db	481	TCTAGCAACAGGGGGGTTTCAACAGCCGAGCTAGGATATGACAGAGTAACTTACAGTTAT 540
OY	553	CCCAGGTATACCTGGAGAGCTAACCCCATGAGCCAGTCACTGCACCTCCATCCCTCCT 612
Db	541	CCCAGGTATACCTGGAGAGCTAACCCCATGAGCCAGTCACTGCACCTCCATCCCTCCT 600
OY	613	ACCAGCTATTCCTCTTACACAGCCGACTATGATATGCAGAGCTTACTCTCAGCAAGAC 672
Db	601	ACCAGCTATTCCTCTTACACAGCCGACTATGATATGCAGAGCTTACTCTCAGCAAGAC 660
OY	673	ACCTATGGGCAACCGAGAGCTATGGAAGAGAGTACTATGCTGCAACAAAGAGCTAT 732
Db	661	ACCTATGGGCAACCGAGAGCTATGGAAGAGAGTACTATGCTGCAACAAAGAGCTAT 720
OY	733	GGGAGAGAGCTCCCACTAGTTAACCAACCCAACTGTGCTCTACAGCCAGCTCCAGT 792
Db	721	GGGAGAGAGCTCCCACTAGTTAACCAACCCAACTGTGCTCTACAGCCAGCTCCAGT 780



[illegible]

Qy	1873	GGAAGACGAGGTGGCCCTGGGGGGGGCCCCCTGACCTTTGATGAAACAGATGGAGGAGA	1932
Db	1861	GGAAACGAGGTGGCCCTGGGGGGGGCCCCCTGACCTTTGATGAAACAGATGGAGGAGA	1920
Qy	1933	AGAGGAGGACCTGAGAGGACCTGGAATAAGTGACGACACCGTCAGAGCGCAGA	1992
Db	1921	AGAGGAGGACCTGAGAGGACCTGGAATAAGTGATTAAGCGAGCACCGTCAGAGCGCAGA	1980
Qy	1993	GATCGCCCTTACTAGATGCAGAGACCCCGCAGA	2025
Db	1981	GATCGCCCTTACTAGATGCAGAGACCCCGCAGA	2013

RESULT 8

ID	AAS62262 standard; cDNA; 2176 BP
1	1
2	2
3	3
4	4
5	5
6	6
7	7
8	8
9	9
10	10
11	11
12	12
13	13
14	14
15	15
16	16
17	17
18	18
19	19
20	20
21	21
22	22
23	23
24	24
25	25
26	26
27	27
28	28
29	29
30	30
31	31
32	32
33	33
34	34
35	35
36	36
37	37
38	38
39	39
40	40
41	41
42	42
43	43
44	44
45	45
46	46
47	47
48	48
49	49
50	50
51	51
52	52
53	53
54	54
55	55
56	56
57	57
58	58
59	59
60	60
61	61
62	62
63	63
64	64
65	65
66	66
67	67
68	68
69	69
70	70
71	71
72	72
73	73
74	74
75	75
76	76
77	77
78	78
79	79
80	80
81	81
82	82
83	83
84	84
85	85
86	86
87	87
88	88
89	89
90	90
91	91
92	92
93	93
94	94
95	95
96	96
97	97
98	98
99	99
100	100

AC AAS62262

DT 14-FEB-2002 (first entry)

DE cdna sequence #49 encoding novel human secreted protein.

KW Human secreted protein; hyperproliferative disorder; autoimmune disorder;

KW infectious disorder; gene therapy; antimicrobial; hepatotropic;

KW immunosuppressive; antirheumatic; ss.

OS Homo sapiens.

PN WO200177291-A2.

PD 18-OCT-2001.

PF 29-MAR-2001; 2001WO-US010485.  
vxy

PR 06-APR-2000; 2000US-0195604P.

PA (GEMY ) GENETICS INST INC.  
VY

PI Wong GG, Clark HF, Fechtel K, Agostino MJ, Howes SH, Resnick RJ;  
PI Culbertson V, Graham TD.

XX  
DB  
WPT : 2003-010900/01

XX  
RE  
New settlement

PT asthma, HIV and Crohn's disease.

PS Claim 1; Page 100; 391pp; English.

CC The present invention relates to the isolation of novel cDNA sequences  
CC which encode human secreted proteins. The cDNA sequences have been

CC derived from a variety of human tissues. The invention also provides a method for producing proteins from these polynucleotide sequences.

CC proteins are useful for identifying compounds that modulate their activity and production and the cell is also useful for identifying

CC compounds that modulate expression of the polynucleotide sequences encoding the secreted proteins. The sequences of the invention are

CC for treating diseases such as hyperproliferative disorders (e.g. car-

CC (SCID), autoimmune disorders (e.g. multiple sclerosis), blood disorders (e.g. thrombocytopenia), inflammatory disorders (e.g. arthritis) and

CC infectious disorders (e.g. hepatitis). The polynucleotide sequences  
CC the invention are also useful in gene therapy. AAS62314-AAS62838

CC represent the cDNA sequences of the invention that encode for novel human secreted proteins

Sequence 2176 Rp. 413 A. 617 C. 563 G. 583 T. 0 U. 0 Other: 50

Query Match 98 1%. Score 1985 4. DB 5. Length 2176.

Best Local Similarity 99.9%; Pred. No. 0;  
Matches 1998: Conservative 0; Mismatches 1; Indels 1;

QY 26 AAGAGAGAAATGCGCTCCACGGATTACGTAAGCTTACGAGCTGCGAGGAGG 85  
DB 2171 AAGGTGAGAAATGCGCTCCACGGATTACGTAAGCTTACGAGCTGCGAGGAGG 2112  
QY 86 GCTACAGTGGCTTACACCGCCAGCCCACTCAAGGATATGCAAGACCAACCGAGCATATG 145  
DB 2111 GCTACAGTGGCTTACACCGCCAGCCCACTCAAGGATATGCAAGACCAACCGAGCATATG 2052  
QY 146 GGCACAAAGCTATGGAAGCTTATGGAAGACCCCACTGATGTCAGTATACCAAGGCTTACA 205  
DB 2051 GGCACAAAGCTATGGAAGCTTATGGAAGACCCCACTGATGTCAGTATACCAAGGCTTACA 1992  
QY 206 CCACTGACCTATGAGGAGAGCGGCTATGCACTTCTTATGAGACGCTCCCACTGCTT 265  
DB 1991 CCACTGACCTATGAGGAGAGCGGCTATGCACTTCTTATGAGACGCTCCCACTGCTT 1932  
QY 266 ATACTACTCCACTGCCCCCGGCAATACAGCCAGCCTGTCAGAGGGTATGCACTGCTG 325  
DB 1931 ATACTACTCCACTGCCCCCGGCAATACAGCCAGCCTGTCAGAGGGTATGCACTGCTG 1872  
QY 326 CTTATGATACCACTGCTGCTACAGTACCAACACCGGCTCTTATGAGCTTCACTGCTG 385  
DB 1871 CTTATGATACCACTGCTGCTACAGTACCAACACCGGCTCTTATGAGCTTCACTGCTG 1812  
QY 386 CATATGAGCACTGAGCTGCTTATCAGGCTATGAGGAGAGCCAGAGCCACTGCACTTA 445  
DB 1811 CATATGAGCACTGAGCTGCTTATCAGGCTATGAGGAGAGCCAGAGCCACTGCACTTA 1752  
QY 446 CAAGACCGCAGATGAGAAACAGGCCCACTGAGTACTGCACTCAATCTTACGACAGGG 505  
DB 1751 CAAGACCGCA - GATGAGAAACAGGCCCACTGAGTACTGCACTCAATCTTACGACAGGG 1693  
QY 506 GTTACAAACAGGCCCACTGAGTATGAGAGAGATTAACAGTATACCCAGGTAAGTCTG 565  
DB 1692 GTTACAAACAGGCCCACTGAGTATGAGAGAGATTAACAGTATACCCAGGTAAGTCTG 1633  
QY 566 GGAAGCTACCCGATGAGAGTCACTGCACTTCACTGCTTACCTACAGGATATCT 625  
DB 1632 GGAAGCTACCCGATGAGAGTCACTGCACTTCACTGCTTACCTTACAGGATATCT 1573  
QY 626 CTACACAGCCGATGATGATGATGATGAGAGAGTACTCTACAGAGAACCTTATGAGGAG 685  
DB 1572 CTACACAGCCGATGATGATGATGATGAGAGAGTACTCTACAGAGAACCTTATGAGGAG 1513  
QY 686 CGAGCACTATGAGACAGAGAGTATGATGATCAACAGAGAGTATGAGGAGAGGCTG 745  
DB 1512 CGAGCACTATGAGACAGAGAGTATGATGATCAACAGAGAGTATGAGGAGAGGCTG 1453  
QY 746 CCACTAGTACCCAGCCCAACCTGATCTTACAGAGCAAGCTCAAGTCAATATAGCCAA 805  
DB 1452 CCACTAGTACCCAGCCCAACCTGATCTTACAGAGCAAGCTCAAGTCAATATAGCCAA 1393  
QY 806 AGAGCAGACGCTAGGAGAGAGATTCATTCGAGAGAGACACCCAGTATGAGGAG 865  
DB 1392 AGAGCAGACGCTAGGAGAGAGATTCATTCGAGAGAGACACCCAGTATGAGGAG 1333  
QY 866 TTTATGAGGAGAGATTCGAGAGATTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 925  
DB 1332 TTTATGAGGAGAGATTCGAGAGATTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1273  
QY 926 CTGATTAACCGGAG 985  
DB 1272 CTGATTAACCGGAG 1213  
QY 986 GGGAG 1045  
DB 1212 GGGAG 1153  
QY 1046 GTGAGACCATGATGAG 1105  
DB 1152 GTGAGACCATGATGAG 1093  
QY 1106 ACTCTGACAAAGTGCATTTATGATGACAAAGATTAATGACAGTGTAGCTCTAGATGATC 1165

DB 1092 ACTCTGACAAAGTGCATTTATGATGACAAAGATTAATGACAGTGTAGCTCTAGATGATC 1033  
QY 1166 TGGCAGACTCTTTTAAGCAGTGTGGGCTTTTAAGTGAACAGAGAGAGAGAGAGAGAG 1225  
DB 1032 TGGCAGACTCTTTTAAGCAGTGTGGGCTTTTAAGTGAACAGAGAGAGAGAGAGAGAG 973  
QY 1226 TGAATCAACTTAAAGCTTCCCTTGTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1285  
DB 972 TGAATCAACTTAAAGCTTCCCTTGTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 913  
QY 1286 ATGAG 1345  
DB 912 ATGAG 853  
QY 1346 GAG 1405  
DB 852 GAG 793  
QY 1406 GTCTGACACCGGAG 1465  
DB 792 GTCTGACACCGGAG 733  
QY 1466 CAG 1525  
DB 732 CAG 673  
QY 1526 TCCCTCCAG 1585  
DB 672 TCCCTCCAG 613  
QY 1586 ACCGAGCTGAG 1645  
DB 612 ACCGAGCTGAG 553  
QY 1646 GAAAG 1705  
DB 552 GAAAG 493  
QY 1706 CGCCCCCGGAG 1765  
DB 492 CGCCCCCGGAG 433  
QY 1766 GCCTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1825  
DB 432 GCCTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 373  
QY 1826 GTGGCTTCCGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1885  
DB 372 GTGGCTTCCGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 313  
QY 1886 GCCCTGG 1945  
DB 312 GCCCTGG 253  
QY 1946 GAGGAGCTGAG 2005  
DB 252 GAGGAGCTGAG 193  
QY 2006 AGATGAG 2025  
DB 192 AGATGAG 173

RESULT 9  
AAS70647  
ID AAS70647 standard; cDNA; 2388 BP.  
XX AAS70647;  
AC AAS70647;  
XX  
XX 13-FEB-2002 (first entry)  
XX  
DB DNA encoding novel human diagnostic protein #6451.



Db	1339	GTGGAATGTTTGAATGGAAAGATTTTCAAGGGAGCAAACTTAAAGTCTCCCTTGCTCGG	1398
Qy	1375	AAGAAAGCTCCATATGAACATATATGGGGGTGGTCTGCCACCCCGTAGGGCAGAGCATG	1434
Db	1399	AAGAAGCTCCAAATGAACATATATGGGGGTGGTCTGCCACCCCGTAGGGCAGAGCATG	1458
Qy	1435	CCACACCACTCCGTGGAGGTCACAGAGAGCCCAAGAGGTCTCGGGGGAGCCCATGGGTGCG	1494
Db	1459	CCACCACTACCTCCGTGGAGGTCACAGAGAGCCCAAGAGGTCTCGGGGGAGCCCATGGGTGCG	1518
Qy	1495	ATGGAGAGCCGTGGAGGATATAGAGAGTCTTCCCTCCAMAGAGACCCCGGGTTCGCCGA	1554
Db	1519	ATGGAGAGCCGTGGAGGATATAGAGAGTCTTCCCTCCAMAGAGACCCCGGGTTCGCCGA	1578
Qy	1555	GGGAAACCCCTTGTGAGAGAGAAACGTCACAGACCGAGCTGGAGACTGGCACTGTCCTCAAT	1614
Db	1579	GGGAAACCCCTTGTGAGAGAGAAACGTCACAGACCGAGCTGGAGACTGGCACTGTCCTCAAT	1638
Qy	1615	CCGGGTGTGAAACCAAGAACTTCGCGCTGGAGAACAGAGTCAACCAATGTAAAGCCCCA	1674
Db	1639	CCGGGTGTGAAACCAAGAACTTCGCGCTGGAGAACAGAGTCAACCAATGTAAAGCCCCA	1698
Qy	1675	AAGCTTAAGGCTTCTCCCGCACCTTTCGCGCCCGGGGTGGATCGTGGCAGAGGT	1734
Db	1699	AAGCTTAAGGCTTCTCCCGCACCTTTCGCGCCCGGGGTGGATCGTGGCAGAGGT	1758
Qy	1735	GGCCCTGGTGCAATCGGGAGAGAAAGAGTGGCTCATGATCGTGTGTCTCCCGGTGGA	1794
Db	1759	GGCCCTGGTGCAATCGGGAGAGAAAGAGTGGCTCATGATCGTGTGTCTCCCGGTGGA	1818
Qy	1795	ATGTTCAAGATGGTCCGCTGGTGAGAACAGAGTGGCTTCGATGGTGGCCGGGGCAATGAC	1854
Db	1819	ATGTTCAAGATGGTCCGCTGGTGAGAACAGAGTGGCTTCGATGGTGGCCGGGGCAATGAC	1878
Qy	1855	CGAGGTGGCTTGTGTGGAGAGAAAGCAGAGTGGCCCTGGGGGGGCCCTCTGACCTTTGATG	1914
Db	1879	CGAGGTGGCTTGTGTGGAGAGAAAGCAGAGTGGCCCTGGGGGGGCCCTCTGACCTTTGATG	1938
Qy	1915	GAACAGATGGAGAGAAAGAGAGAGACGTGGAGGACCTGGAAAAATGATTAAGCGCAG	1974
Db	1939	GAACAGATGGAGAGAAAGAGAGAGACGTGGAGGACCTGGAAAAATGATTAAGCGCAG	1998
Qy	1975	CACCGTCAGAGACGCAAGATGGCGCTTAATGATGCAAGACCCCGCAGA	2025
Db	1999	CACCGTCAGAGACGCAAGATGGCGCTTAATGATGCAAGACCCCGCAGA	2049

RESULT 10  
ACN37537  
ID ACN37537 standard; cDNA; 2177 BP.  
ACN37537;  
DT 18-NOV-2004 (first entry)  
XX Tumour-associated antigenic target (TAT) cDNA DNA33971, SEQ ID NO:490.  
XX  
XX  
XX Tumour-associated antigenic target; TAT; human; overexpression; cancer;  
XX  
XX  
XX Tumour diagnosis; cell proliferative disorder; breast cancer;  
XX  
XX colorectal cancer; lung cancer; ovarian cancer; liver cancer;  
XX  
XX central nervous system cancer; bladder cancer; pancreatic cancer;  
XX  
XX cervical cancer; melanoma; leukaemia; hybridisation probe;  
XX  
XX chromosome identification; chromosome mapping; gene mapping;  
XX  
XX gene therapy; cytostatic; gene; ss.  
OS Homo sapiens.  
XX  
XX  
XX WO2004030615-A2.  
XX  
XX  
XX 15-APR-2004.  
XX  
XX  
XX 29-SEP-2003; 2003WO-US028547.

```

PR 02-0CT-2002; 2002US-0414971P.
XX
XX (GETH ) GENENTECH INC.
XX
XX Wu TD, Zhang Z, Zhou Y;
XX
XX MPI; 2004-347921/32.
XX
XX P-PSDB; ABM80197.
XX
XX New tumor-associated antigenic target polypeptides and nucleic acids,
XX useful in preparing a medicament for treating or detecting a
XX proliferative disorder, e.g. breast, lung, colorectal, ovarian or
XX prostate cancer or tumor.
XX
XX Claim 1; SEQ ID NO 490; 7273bp; English.
XX
XX The invention relates to human tumor-associated antigenic target (TAT)
XX polypeptides, and their related nucleic acids. The TAT polypeptides are
XX overexpressed in cancer tissues compared to normal tissues, and may thus
XX serve as effective targets for the diagnosis and treatment of cancer in
XX mammals. The invention also relates to nucleic acid and polypeptide
XX sequences at least 80% identical to the TAT nucleic acids and
XX polypeptides; expression vectors and host cells comprising a TAT nucleic
XX acid; an antibody specific for a TAT polypeptide; a peptide or organic
XX molecule which binds to a TAT polypeptide; fusion proteins comprising a
XX TAT polypeptide; and methods and compositions for the treatment or
XX diagnosis of cancer in mammals. TAT polypeptides, nucleic acids,
XX antibodies, antagonists, binding molecules and compositions are useful
XX for diagnosing or treating a cell proliferative disorder associated with
XX increased TAT expression, particularly cancers such as breast cancer,
XX colorectal cancer, lung cancer, ovarian cancer, liver cancer, bladder
XX cancer, pancreatic cancer, cervical cancer, cancers of the central
XX nervous system, melanoma and leukaemia. TAT nucleic acids may further be
XX used as hybridisation probes, in chromosome and gene mapping, in
XX chromosome identification and in gene therapy. The present sequence
XX represents a TAT nucleic acid of the invention
XX
SQ Sequence 2177 BP; 602 A; 539 C; 605 G; 431 T; 0 U; 0 Other;
XX
Query Match 92.8%; Score 1879.6; DB 13; Length 2177;
Best Local Similarity 96.2%; Pred. No. 0;
Matches 1946; Conservative 0; Mismatches 74; Indels 2; Gaps 2;
XX
4 GCAGCTTGAGGAAACGAGAGGAAAGAGAGAAATGGCGTCCACAGGATTACAGTACCAT 63
Db 1 GGAGCTTGAGGAAACGAGAGGAAAGAGAGAAATGGCGTCCACAGGATTACAGTACCAT 60
QY AGCCAGAGCTGAGGCGGAGCAGGAGGCTACAGTGTCTTACACCGCCAGCCACTCAAGATAT 123
Db 61 AGCCAGAGCTGAGGCGGAGCAGGAGGCTACAGTGTCTTACACACCCAGCCACTCAAGATAT 120
QY 124 GCACAGACCAACCCAGGCATATATGGCCAAACAAAGCTATAGGACCTTATGACAGCCCACTGAT 183
Db 121 GCACAGACCAACCCAGGCATATATGGCCAAACAAAGCTATAGGACCTTATGACAGCCCACTGAT 180
QY 184 GTCAAGCTATACCAAGGCTCAGACCACTGACAACTATGCGGCAAGCCGCTTATGCAACTTCT 243
Db 181 GTCAAGCTATACCAAGGCTCAGACCACTGACAACTATGCGGCAAGCCGCTTATGCAACTTCT 240
QY 244 TATGAGACGCTCCCACTGCTTATATCTACCTCAACTGCCCCCAAGGGCATACAGCCAGCCT 303
Db 241 TATGAGACGCTCCCACTGCTTATATCTACCTCAACTGCCCCCAAGGGCATACAGCCAGCCT 300
QY 304 GTCCAGAGGGATATGGCATCTGATGCTTATGATATACCAACTGCTACAGTACCAACCAACCCAG 363
Db 301 GTCCAGAGGGATATGGCATCTGATGCTTATGATATACCAACTGCTACAGTACCAACCAACCCAG 360
QY 364 GCGTCTTATGACGCTCACTGATGCTATGATGCACTCAGCCTGCTTATTCAGCCTTATGGGCGAG 423
Db 361 GCGTCTTATGACGCTCACTGATGCTATGATGCACTCAGCCTGCTTATTCAGCCTTATGGGCGAG 420
QY 424 CAGCCAGACGCACTGCACTTACCAAGACCGCAGATGGAACCAAGCCCACTGACACTGAT 483

```

Db 421 CAGCCAGTACGCTGACCTTACAGACTGACAGATGAAACAGCCCTGAGACTAGT 480  
Qy 484 CAACTTCAATCTAGCAGAGGGGTTTAAACAGCCAGCTAGAGATATGACAGAGTAC 543  
Db 481 CAACCTCAATCTAGCAGAGGGGTTTAAACAGCCAGCTAGAGATATGACAGAGTAC 540  
Qy 544 TACAGTTATCTCCAGAGTACTGAGAGTACCCATGACAGCCAGTCACTTCCATCC 603  
Db 541 TGAAGTTATCTCCAGAGTACTGAGAGTACCCATGACAGCCAGTCACTTCCATCC 600  
Qy 604 TACCTCTCTACAGACTATCTCTACAGCCAGCTAGTTATGATCAGAGAGTACTCT 663  
Db 601 TACCTCTCTACAGACTATCTCTACAGCCAGCTAGTTATGATCAGAGAGTACTCT 660  
Qy 664 CAGCAGAACACTATGAGGCAACCGAGAGCTATGACAGAGTATGCTATGCTACAA 723  
Db 661 CAGCAGAACACTATGAGGCAACCGAGAGCTATGACAGAGTATGCTATGCTACAA 720  
Qy 724 AGCAGTATGAGGCAAGCCTCCCACTAGTTACCCCAACTGATCTTACAGCCAA 783  
Db 721 AGCAGTATGAGGCAAGCCTCCCACTAGTTACCCCAACTGATCTTACAGCCAA 780  
Qy 784 GCTCAAGTCAATATGAGCAACAGAGAGCTAGAGGAGAGAGTTCATTCGAGCAG 843  
Db 781 GCTCAAGTCAATATGAGCAACAGAGAGCTAGAGGAGAGAGTTCATTCGAGCAG 840  
Qy 844 GACCACTCCAGTATGAGTGTGTTTATGAGGCAAGTCTGAGAGATTTTCGAGCAG 903  
Db 841 GACCACTCCAGTATGAGTGTGTTTATGAGGCAAGTCTGAGAGATTTTCGAGCAG 900  
Qy 904 GAGAACCGAGACATGAGTGCCTGATTAACCGGGGCAAGGGGAGAGGAGATTTGATCT 963  
Db 901 GAGAACCGAGACATGAGTGCCTGATTAACCGGGGCAAGGGGAGAGGAGATTTGATCT 960  
Qy 964 GAGAGCATGAGAGAGTGGGGGGGAGAGAGAGCGGGTGAATGGGCAAGCGCTGAGAG 1023  
Db 961 GAGAGCATGAGAGAGTGGGGGGGAGAGAGAGCGGGTGAATGGGCAAGCGCTGAGAG 1020  
Qy 1024 CGAGTGGCTTCAATAGCCTGTGTGACCCATGATAGAGACCAAGATCTTGAATCTAGGC 1083  
Db 1021 CAGTTGGCTTCAATAGCCTGTGTGACCCATGATAGAGACCAAGATCTTGAATCTAGGC 1080  
Qy 1084 CCACCTGTATATCCAGATGAAAGACTCTGACAAAGTGCATTTATGTATCAAGATTAAT 1143  
Db 1081 CCACCTGTATATCCAGATGAAAGACTCTGACAAAGTGCATTTATGTATCAAGATTAAT 1140  
Qy 1144 GACAGTGTGACTAGATGATCTGGGAGACTCTTTAAGAGAGTGGGGTTGTTAAGATG 1203  
Db 1141 GACAGTGTGACTAGATGATCTGGGAGACTCTTTAAGAGAGTGGGGTTGTTAAGATG 1200  
Qy 1204 AACAGAGAACTGGGCAACCATATGATCAGATCTTACCTGAGCAAGAAAGAGAGCC 1263  
Db 1201 AACAGAGAACTGGGCAACCATATGATCAGATCTTACCTGAGCAAGAAAGAGAGCC 1260  
Qy 1264 AAAGGCGATGCCAGTGTCTTATGAAAGCCCACTGCAAGGCTGCTGAGAAATGG 1323  
Db 1261 AAAGGCGATGCCAGTGTCTTATGAAAGCCCACTGCAAGGCTGCTGAGAAATGG 1320  
Qy 1324 TTTGATGGGAAAGATTTTCAAGGGAGAGAACTTAAATCTCCCTTGTGCGAAGAGAGCT 1383  
Db 1321 TTTGATGGGAAAGATTTTCAAGGGAGAGAACTTAAATCTCTTGTGCGAAGAGAGCT 1380  
Qy 1384 CCAATGAAACAGTATGCGAGGTGTGTGCAACCCGATGAGGAGAGAGCTCCACCA 1443  
Db 1381 CCAATGAAACAGTATGCGAGGTGTGTGCAACCCGATGAGGAGAGAGCTCCACCA 1440  
Qy 1444 CTCCTGTGAGGTCCAGAGAGGCCCAAGAGTCTTGGGAGACCCATGAGTGGAGGC 1503  
Db 1441 CTCCTGTGAGGTCCAGAGAGGCCCAAGAGTCTTGGGAGACCCATGAGTGGAGGC 1500  
Qy 1504 CGTGAAGAGATGAGAGAGCTTCCCTCAAGAGAGACCCCGGGTTCCGAGGAAACCC 1563

Db 1501 CGTGAAGAGATGAGAGAGCTTCCCTCAAGAGAGACCCAGGGTTCCGAGGAAACCC 1560  
Qy 1564 TCTGAGAGAGAGAAAGCTTCAGCAACGAGCTGAGAGACTGAGAGTGTCCAAATCCGGTGT 1623  
Db 1561 TCTGAGAGAGAGAAAGCTTCAGCAACGAGCTGAGAGAGAGAGTGTCCAAATCCGGTGT 1620  
Qy 1624 GGAACCAAGAACTTCCGCTGAGAGAGAGAGTGAACAGAGTAAAGGCCCAAGAGCTGAA 1683  
Db 1621 GGAACCAAGAACTTCCGCTGAGAGAGAGAGTGAACAGAGTAAAGGCCCAAGAGCTGAA 1680  
Qy 1684 GGCCTCTCCCGCACCCTTCCGCGCCCGGGGTGATGATGATGAGAGAGTGCCTGTGT 1743  
Db 1681 GGCCTCTCCCGCACCCTTCCGCGCGGGGTGATGATGATGAGAGAGTGCCTGTGT 1740  
Qy 1744 GGCATGCGGGAGAGAAAGAGTGCCTTCATGATGATGATGATGATGATGATGATGAT 1803  
Db 1741 GGCATGCGGGAGAGAAAGAGTGCCTTCATGATGATGATGATGATGATGATGATGAT 1800  
Qy 1804 GGTGGCTGTGTGTGAGACAGAGTGTGCTTCCGTGTGTGCGGGGCAATGACCGAGTGGC 1863  
Db 1801 GGTGGCTGTGTGTGAGACAGAGTGTGCTTCTGTGTGTG-CTGGCATGAGACCGAGTGGC 1859  
Qy 1864 TTTGTGTGAGAGAGAGAGAGTGCCTTCGAGAGAGCCCTTGAATGATGAGACAGATG 1923  
Db 1860 TTTGTGTGAGAGAGAGAGAGTGCCTTCGAGAGAGCCCTTGAATGATGATGAGACAGATG 1918  
Qy 1924 GAGAGAAAG 1983  
Db 1919 GAGAGAAAG 1978  
Qy 1984 GAGCGCAGAGATGCGCCCTTCTAGATGACAGAGACCCGCGAGA 2025  
Db 1979 GAGCGCAGAGATGCGCCCTTCTAGATGACAGAGAGAGAGAGAGAGAGAGAGAGAG 2020

RESULT 11  
AAS6263/c  
ID AAS6263 standard; cDNA; 2273 BP.  
XX  
AC AAS62623;  
XX  
DT 14-FEB-2002 (first entry)  
XX  
XX  
DE cDNA sequence #410 encoding novel human secreted protein.  
XX  
XX  
KW Human secreted protein; hyperproliferative disorder; autoimmune disorder;  
KW immune deficiency disorder; blood disorder; inflammatory disorder;  
KW infectious disorder; gene therapy; antimicrobial; hepatotropic;  
KW immunosuppressive; antineumatic; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200177291-A2.  
XX  
PD 18-OCT-2001.  
XX  
XX 23-MAR-2001; 2001WO-US010485.  
PF 23-MAR-2001; 2001WO-US010485.  
XX  
PR 06-APR-2000; 2000US-0195604P.  
XX  
XX (GENE) GENETICS INST INC.  
XX  
XX Wong GG, Clark HF, Recheil K, Agostino MJ, Howes SH, Resnick RJ;  
PI Gulukota K, Graham JR;  
XX  
XX DR WPI; 2002-010900/01.  
XX  
XX  
XX New polynucleotides encoding secreted proteins useful for treating e.g.  
PT asthma, HIV and Crohn's disease.  
XX  
XX  
PS Claim 1; Page 291; 391pp; English.  
XX  
XX The present invention relates to the isolation of novel cDNA sequences



CC which encode human secreted proteins. The cDNA sequences have been  
CC derived from a variety of human tissues. The invention also provides a  
CC method for producing proteins from these polynucleotide sequences. The  
CC proteins are useful for identifying compounds that modulate their  
CC activity and production, and the cell is also useful for identifying  
CC compounds that modulate expression of the polynucleotide sequences  
CC encoding the secreted proteins. The sequences of the invention are useful  
CC for treating diseases such as hyperproliferative disorders (e.g. cancer),  
CC immune deficiency disorders (e.g. severe combined immunodeficiency  
CC (SCID)), autoimmune disorders (e.g. multiple sclerosis), blood disorders  
CC (e.g. thrombocytopenia), inflammatory disorders (e.g. arthritis) and  
CC infectious disorders (e.g. hepatitis). The polynucleotide sequences of  
CC the invention are also useful in gene therapy. AA562214-AA56238  
CC represent the cDNA sequences of the invention that encode for novel human  
CC secreted proteins  
XX  
SQ Sequence 2273 BP; 477 A; 636 C; 562 G; 598 T; 0 U; 0 Other;  
Query Match 91.7%; Score 1856.6; DB 6; Length 2273;  
Best Local Similarity 97.1%; Pred. No. 0;  
Matches 1933; Conservative 0; Mismatches 4; Indels 54; Gaps 2;  
QY 35 AAATGGCGTCCAGGATTAGTACTTATAGCCAGCTGACGCGACAGAGGCTTACAGTG 94  
DB 2268 AAGGTGCTCCAGGATTACGATCTATAGCCAGCTGACGCGACAGAGGCTTACAGTG 2209  
QY 95 CTTACACCGCCGACGCTCAGAGATATGACAGACACCGGCTATGGGCAACAA 154  
DB 2208 CTTACACCGCCGACGCTCAGAGATATGACAGACACCGGCTATGGGCAACAA 2149  
QY 155 GCTATGGAACCTATGACAGAGCCCATGATGACGCTATACCCAGGCTCAGACCACTGCA 214  
DB 2148 GCTATGGAACCTATGACAGAGCCCATGATGACGCTATACCCAGGCTCAGACCACTGCA 2089  
QY 215 CCTATGGGACAGACCCGCTATGCACTTCTATGAGACGCTCCCACTGTTATATCTACTC 274  
DB 2088 CCTATGGGACAGACCCGCTATGCACTTCTATGAGACGCTCCCACTGTTATATCTACTC 2029  
QY 275 CAATGCGCCCGACGCTATGACAGAGCCGCTGCGAGGGGTATGGACCTGGTCTTATATATA 334  
DB 2028 CAATGCGCCCGACGCTATGACAGAGCCGCTGCGAGGGGTATGGACCTGGTCTTATATATA 1969  
QY 335 CCAACACCTGCTACAGTACCAACCAACGAGGCTCTTATGAGCTCAGTCTGATATGCA 394  
DB 1968 CCAACACCTGCTACAGTACCAACCAACGAGGCTCTTATGAGCTCAGTCTGATATGCA 1909  
QY 395 CTGAGCGCTGCTTATCAGGCTTATGAGGACGACGAGGCTCCTTACAAAGACCGC 454  
DB 1908 CTGAGCGCTGCTTATCAGGCTTATGAGGACGACGAGGCTCCTTACAAAGACCGC 1849  
QY 455 AGATGGAACCAAGCCCACTGAGACTGACCTCAATCTAGACAGGGGGTTTACAAAC 514  
DB 1848 AGATGGAACCAAGCCCACTGAGACTGACCTCAATCTAGACAGGGGGTTTACAAAC 1789  
QY 515 AGCCAGCCTAGATATGACAGAGTAACTACGTTATCCAGGTAAGTCTGGAGTACC 574  
DB 1788 AGCCAGCCTAGATATGACAGAGTAACTACGTTATCCAGGTAAGTCTGGAGTACC 1729  
QY 575 CCATGAGCCAGTACGACCTTCATCTACCTCTTACAGGTAATCTTACAGAGC 634  
DB 1728 CCATGAGCCAGTACGACCTTCATCTACCTCTTACAGGTAATCTTACAGAGC 1669  
QY 635 CGATCTATGATCAAGAGCTTACTCTGAGCAACCACTATGGGCAACCGAGAGCT 694  
DB 1668 CGATCTATGATCAAGAGCTTACTCTGAGCAACCACTATGGGCAACCGAGAGCT 1609  
QY 695 ATGAGCAGAGAGTATGATGATGATCAACAAAGAGCTATGGGACAGGCTCCACTAGTT 754  
DB 1608 ATGAGCAGAGAGTATGATGATGATCAACAAAGAGCTATGGGACAGGCTCCACTAGTT 1549  
QY 755 ACCCAGCCCAATCTGATCTTACAGCCAAAGCTTCAAGTCAATATAGCCAAACAGAGCA 814  
DB 1548 ACCCAGCCCAATCTGATCTTACAGCCAAAGCTTCAAGTCAATATAGCCAAACAGAGCA 1489

QY 815 GCTACGGGAGAGAGTTCATTCCGACAGACCAACCCAGTACATGGGTTTATGGGC 874  
DB 1488 GCTACGGGAGAGAGTTCATTCCGACAGACCAACCCAGTACATGGGTTTATGGGC 1429  
QY 875 AGGAGCTCTGAGAGATTTTCCGACAGAGAGAACCCGAGCATGAGTGGCTCTGATACC 934  
DB 1428 AGGAGCTCTGAGAGATTTTCCGACAGAGAGAACCCGAGCATGAGTGGCTCTGATACC 1369  
QY 935 GGGGACGGGGAAGAGGGGGATTTGATCTGAGAGGCAATGACAGAGTGGGGGGGAGAG 994  
DB 1368 GGGGACGGGGAAGAGGGGGATTTGATCTGAGAGGCAATGACAGAGTGGGGGGGAGAG 1309  
QY 995 GACGGGATGAAATGGGACGCTGAGAGAGAGTGGCTTCAATTAAGCTGTGACCCA 1054  
DB 1308 GACGGGATGAAATGGGACGCTGAGAGAGAGTGGCTTCAATTAAGCTGTGACCCA 1252  
QY 1055 TGGATGAGAGACCAAGATCTTATCTAGGCCCACCTGTAGATCCAGATGAACCTTGACA 1114  
DB 1251 TGGATGAGAGACCAAGATCTTATCTAGGCCCACCTGTAGATCCAGATGAACCTTGACA 1192  
QY 1115 ACAGTCAATTTTATGTAACAAGATTTAAATGACAGTGTGACTCTAGATGATCTGGACACT 1174  
DB 1191 ACAGTCAATTTTATGTAACAAGATTTAAATGACAGTGTGACTCTAGATGATCTGGACACT 1132  
QY 1175 TCTTTAAGAGTGGGTTGTTAAGATGAAACAAGAACTGGGCAACCAATGATCCACA 1234  
DB 1131 TCTTTAAGAGTGGGTTGTTAAGATGAAACAAGAACTGGGCAACCAATGATCCACA 1072  
QY 1235 TCTACCTGGAACAAGAAACAAGAAAGCCAAAGGCGATGCAAGTGTCTATGAAGACC 1294  
DB 1071 TCTACCTGGAACAAGAAACAAGAAAGCCAAAGGCGATGCAAGTGTCTATGAAGACC 1012  
QY 1295 CACCCACTGCGCAAGGCTGCGGCTGAGATGTTTGAATGGGAAAGTTTCAAGGGAGCAAC 1354  
DB 1011 CACCCACTGCGCAAGGCTGCGGCTGAGATGTTTGAATGGGAAAGTTTCAAGGGAGCAAC 952  
QY 1355 TTAAGTCTCCCTTGTGCGAAGAACTCTCAATGAAACGATATCGGGGTGTCTGCGAC 1414  
DB 951 TTAAGTCTCCCTTGTGCGAAGAACTCTCAATGAAACGATATCGGGGTGTCTGCGAC 892  
QY 1415 CCGGTGAGGACAGAGCATGACCAACCACTCCGTGAGGTCCAGAGGCGCCAGAGGTC 1474  
DB 891 CCGGTGAGGACAGAGCATGACCAACCACTCCGTGAGGTCCAGAGGCGCCAGAGGTC 832  
QY 1475 CTGGGGGACCCATGGGTCCGATGGGAGCCGTGAGAGATGAGAGAGGCTTCCCTCA 1534  
DB 831 CTGGGGGACCCATGGGTCCGATGGGAGCCGTGAGAGATGAGAGAGGCTTCCCTCA 772  
QY 1535 GAGGACCCCGGGGTTCCCGAGGGAACCCCTCTGAGAGGAAACGTCACAGACGAGCTG 1594  
DB 771 GAGGACCCCGGGGTTCCCGAGGGAACCCCTCTGAGAGGAAACGTCACAGACGAGCTG 712  
QY 1595 GAGACTGGCAGTGTCCCAATCCGGGTTTGGAAACAAGAACTTCCCTGAGAAACAGAGT 1654  
DB 711 GAGACTGGCAGTGTCCCAATCCGGGTTTGGAAACAAGAACTTCCCTGAGAAACAGAGT 652  
QY 1655 GCAACAGTGTAAAGGCCCAACAGCTGAGAGCTTCCCTCCGCAACCTTCCCGCCCCG 1714  
DB 651 GCAACAGT----- 643  
QY 1715 GTGGTATCGTGGCAGAGTGGCCCTGTTGGCAATGCGGGAGGAAAGAGTGGCTCATGG 1774  
DB 642 GTGGTATCGTGGCAGAGTGGCCCTGTTGGCAATGCGGGAGGAAAGAGTGGCTCATGG 583  
QY 1775 ATCGTGTGTGCCCGTGAATGTTCAAGAGTGGCCGTGTGAGAGACAGAGTGGCTTCC 1834  
DB 582 ATCGTGTGTGCCCGTGAATGTTCAAGAGTGGCCGTGTGAGAGACAGAGTGGCTTCC 523  
QY 1835 GTGGTGGCCGGGGCTAGACCGAGTGGCTTGTGTGAGGAAACAAGAGTGGCCCTGGGG 1894  
DB 522 GTGGTGGCCGGGGCTAGACCGAGTGGCTTGTGTGAGGAAACAAGAGTGGCCCTGGGG 463





QY 1141 AATGACAGTGTGACTCTAGATGATCTGGCAAGCTTCTTTAAGAGTGTGGGTTTAAAG 1200  
 DB 1158 AATGACAAATGTGACTCTGATGATCTGGCAAGCTTCTTTAAGAGTGTGGGTTTAAAG 1217  
 QY 1201 ATGAACAAGAGTGTGGGCAACCATGATCTACATCTACCTGGAACAAGAAAGAGAAAG 1260  
 DB 1218 ATGAACAAGAGTGTGGGCAACCATGATCTACATCTACCTGGAACAAGAAAGAGAAAG 1277  
 QY 1261 CCCAAGGCGATGCCAGTGTCTCTATGAAGACCACCTGCGCAAGGTCGCCGTGAA 1320  
 DB 1278 CTTAAAGGGGAGCCCAAGTGTCTCTATGAAGACCACCTGCGCAAGGTCGCCGTGAA 1337  
 QY 1321 TGGTTTGAATGGGAAAGTTTTCAAAGGAGCAACTTAAATCTCCCTGTCTGGAAAG 1380  
 DB 1338 TGGTTTGAATGGGAAAGTTTTCAAAGGAGCAACTTAAATCTCTTCTCCCAAGAG 1397  
 QY 1381 CTTCAATGAACAGTATGCGGGGTGTCTGCAACCCCTGAGGGCAGAGGCAATGCCACA 1440  
 DB 1398 CTTCAATGAACAGTATGCGGGGTGTCTGCAACCCCTGAGGGCAGAGGCAATGCCACA 1457  
 QY 1441 CCACTCCGTGAGGTCCAGAGGCCCAAGAGTCTTGGGGACCCATGSGTCCGATGGGA 1500  
 DB 1458 CCACTCCGTGAGGTCTGAGTGGCCCAAGAGGCCCTGAGAGCAATGSGTCCGATGGGA 1517  
 QY 1501 GGGCTGTGAGAGATTAAGAGAGTCTCCCTCAAGAGAGACCCGGGGTTCGAGAGGAG 1560  
 DB 1518 GGGCTGTGAGAGATTAAGAGAGTCTCCCTCAAGAGAGAGCCCGAGGCTCCAGAGAAC 1577  
 QY 1561 CCTCTGAGAGAGAAAGTTCAGACCCGAGCTGAGAGTGTGCAATCCCGAGT 1620  
 DB 1578 CCTCTGAGAGAGAAAGTTCAGACCCGAGCTGAGAGTGTGCAATCCCGAGT 1637  
 QY 1621 TGTGAAACCAAGAACTTCTGCTGAGAAACAGAGTGAACCAAGTGTAAAGCCCAAGCCT 1680  
 DB 1638 TGTGAAACCAAGAACTTCTGCTGAGAAACAGAGTGAACCAAGTGTAAAGCCCAAGCCT 1697  
 QY 1681 GAAGGCTTCTCCCGCAACCTTCCGCGCCCGGGGTGTGATCTGTGCAAGAGTGGCCCT 1740  
 DB 1698 GAAGGCTTCTCCCGCAACCTTCCGCGCCCGGGGTGTGATCTGTGCAAGAGTGGCCCT 1757  
 QY 1741 GGTGCAATGCGGGAGAGAGAGTGGCTCATGATGATGATGATGATGATGATGATGATGAT 1800  
 DB 1758 GGTGCAATGCGGGAGAGAGAGTGGCTCATGATGATGATGATGATGATGATGATGATGAT 1817  
 QY 1801 AGAGTGTGCGGTGTGAGACAGAGTGTGCTTCCGTGTGTGCGGGCATGAGCCAGGT 1860  
 DB 1818 AGAGTGTGCGGTGTGAGACAGAGTGTGCTTCCGTGTGTGCGGGCATGAGCCAGGT 1877  
 QY 1861 GGTGTTGTGAGAGAAACAGAGTGTGCTTGGGGGGGGCCCTGTGACCTTTGATGAAACAG 1920  
 DB 1878 GGTGTTGTGAGAGAAACAGAGTGTGCTTGGGGGGGGCCCTGTGACCTTTGATGAAACAG 1937  
 QY 1921 ATGGAGAGAGAAAGAGAGAGTGTGAGAGACCTGAAATGATTAAGGCGAGACCGGT 1980  
 DB 1938 ATGGAGAGAGAAAGAGAGAGTGTGAGAGACCTGAAATGATTAAGGCGAGACCGGT 1997  
 QY 1981 CAGAGCGCAGAGATCGGCTCTACTAGA 2008  
 DB 1998 CAGAGCGCAGAGATCGGCTCTACTAGA 2025

## RESULT 13

ADAS3506 ID ADAS3506 standard; cDNA; 1988 BP.

XX ADAS3506;

XX 20-NOV-2003 (first entry).

XX Human coding sequence, SEQ ID 1074.

XX Cytosolic; Anti-inflammatory; Osteoprotective; Neuroprotective; Nootropic;

KM Gene Therapy; human; secretory protein; membrane proteins; cancer;  
 KM Inflammatory disease; osteoporosis; neurological disease; gene; ss.  
 OS Homo sapiens.  
 XX BP1293569-A2.  
 XX 19-MAR-2003.  
 XX 21-MAR-2002; 2002EP-0006586.  
 XX 14-SEP-2001; 2001JP-0032831.  
 PR 24-JAN-2002; 2002US-0350435P.  
 XX (HELI-) HELIX RES INST.  
 PA (REAS-) RES ASSOC BIOTECHNOLOGY.  
 XX Isogai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S,  
 PI Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I,  
 PI Seki N, Yoshikawa T, Otsuka M, Nagahara K, Masuno Y;  
 XX WPI; 2003-395539/38.  
 DR P-PDB; ADAS5145.  
 XX New polynucleotides encoding full-length polypeptides, e.g. secretory  
 PT and/or membrane proteins, useful for developing medicines for diseases in  
 PT which the gene is involved, or as target molecules for gene therapy.  
 PS Claim 1; SEQ ID NO 1074; 205pp; English.  
 CC The present invention relates to novel human secretory or membrane  
 CC proteins (ADAS4072-ADAS5710) and their coding sequences (ADAS433-  
 CC ADAS4071). The coding sequences are useful in the gene therapy of  
 CC diseases caused by abnormalities of the proteins, e.g. cancer,  
 CC inflammatory diseases, osteoporosis or neurological disease.  
 XX Sequence 1988 BP; 519 A; 504 C; 584 G; 381 T; 0 U; 0 Other;  
 SQ

Query Match 81.6%; Score 1652.4; DB 10; Length 1988;  
 Best Local Similarity 91.5%; Pred. No. 0; Mismatches 1; Indels 168; Gaps 1;  
 Matches 1831; Conservative 0;

QY 26 AAGAGAGAAATGAGGTGTCAGAGATTAAGTCACTATGAGCAAGTGCAGCCAGCAGG 85  
 DB 1 AAGAGAGAAATGAGGTGTCAGAGATTAAGTCACTATGAGCAAGTGCAGCCAGCAGG 60  
 QY 86 GCTACAGTGTTCACCGCCAGCCCACTCAGAGATTTGCAAGACCAACCCAGGCATATG 145  
 DB 61 GCTACAGTGTTCACCGCCAGCCCACTCAGAGATTTGCAAGACCAACCCAGGCATATG 120  
 QY 146 GCGAACAAGCTATGAACTTATGACAGCCCACTGATGATGATGATGATGATGATGATGAT 205  
 DB 121 GCGAACAAGCTATGAACTTATGACAGCCCACTGATGATGATGATGATGATGATGATGAT 180  
 QY 206 CCACTGCACTTATGAGCAGACCGCTATGCACTTCTTATGAGCAGCTTCCACTGCTT 265  
 DB 181 CCACTGCACTTATGAGCAGACCGCTATGCACTTCTTATGAGCAGCTTCCACTGCTT 240  
 QY 266 ATACTACTCAACTGCCCGCCAGGCATACAGCCAGCTGTCCAGGGGTATGAGCTGTG 325  
 DB 241 ATACTACTCAACTGCCCGCCAGGCATACAGCCAGCTGTCCAGGGGTATGAGCTGTG 300  
 QY 326 CTTATGATACACAGCTGTCTACAGTCAACCCAGGCGCTCTCTATGAGCTCAGCTG 385  
 DB 301 CTTATGATACACAGCTGTCTACAGTCAACCCAGGCGCTCTCTATGAGCTCAGCTG 360  
 QY 386 CATATGCACTCAGCTGTCTTATGAGCTTATGAGCTTATGAGCTTATGAGCTTATGAGCT 445  
 DB 361 CATATGCACTCAGCTGTCTTATGAGCTTATGAGCTTATGAGCTTATGAGCTTATGAGCT 420  
 QY 446 CAGAGCCGAGATGAGAAACAGCCCACTGAGACTGACTCAACTCAATTAAGCAGGGG 505  
 DB 421 CA----- 422

QY 506 GTTAAACACAGCCAGCTAGATATGAGACAGAGTAAGTAAGTTATCCAGGTAAGCTG 565  
Db 423 ----- 422  
QY 566 GGAAGTACCCATGACAGCAAGTCACTGCACTCCATCTTACCTTACAGCATTTCT 625  
Db 423 -----AGTTATTTCT 432  
QY 626 CTACACAGCCGACTAGTTATGATGACAGCACTTCTCAGCAGAAACAAGTTATGAGCAAC 685  
Db 433 CTACACAGCCGACTAGTTATGATGACAGCACTTCTCAGCAGAAACAAGTTATGAGCAAC 492  
QY 686 CGAGCACTATGAGCAGCAGAGTGTGCTATGCTCAACAAAGAGCTATGAGGAGAGAGCTC 745  
Db 493 CGAGCACTATGAGCAGCAGAGTGTGCTATGCTCAACAAAGAGCTATGAGGAGAGAGCTC 552  
QY 746 CCACTAGTTACCAACCCCAAACTGATCTTACAGCAGCAAGCTCAAGTCAATATATGAGCAAC 805  
Db 553 CCACTAGTTACCAACCCCAAACTGATCTTACAGCAGCAAGCTCAAGTCAATATATGAGCAAC 612  
QY 806 AGAGCAGCAGCTACAGGAGCAGCAGATTCATTCCGACAGACCAACCCAGTATGAGTGTG 865  
Db 613 AGAGCAGCAGCTACAGGAGCAGCAGATTCATTCCGACAGACCAACCCAGTATGAGTGTG 672  
QY 866 TTTATGAGCAGAGATCTGAGAGATTTTCCGACAGAGAGAGACCCGAGCATGAGTGTG 925  
Db 673 TTTATGAGCAGAGATCTGAGAGATTTTCCGACAGAGAGAGACCCGAGCATGAGTGTG 732  
QY 926 CTGATTAACCGGGGAGAGGGGAAAGAGGGGAAATTGATCGTGAAGGAGATGAGCAGAGTGTG 985  
Db 733 CTGATTAACCGGGGAGAGGGGAAAGAGGGGAAATTGATCGTGAAGGAGATGAGCAGAGTGTG 792  
QY 986 GGGAGAGAGAGACGCGGTGAAATGGGCAACGCTGAGAGAGAGAGTGTGCTTCAATTAAGCTG 1045  
Db 793 GGGAGAGAGAGACGCGGTGAAATGGGCAACGCTGAGAGAGAGAGTGTGCTTCAATTAAGCTG 852  
QY 1046 GTGAGCCCATGAGTGAAGAGACCAATCTTGATCTAGGAGCCACTGTATGATCCAGATGAG 1105  
Db 853 GTGAGCCCATGAGTGAAGAGACCAATCTTGATCTAGGAGCCACTGTATGATCCAGATGAG 912  
QY 1106 ACTCTGACACAGTGAATTTATGATGAAGAAATTAAGACAGTGAATCTTATGATGATC 1165  
Db 913 ACTCTGACACAGTGAATTTATGATGAAGAAATTAAGACAGTGAATCTTATGATGATC 972  
QY 1166 TGGCAGACTCTTTAAGCAGTGTGGGTGTTAAGATGAACAGAGAGATGAGGAGCAACCA 1225  
Db 973 TGGCAGACTCTTTAAGCAGTGTGGGTGTTAAGATGAACAGAGAGATGAGGAGCAACCA 1032  
QY 1226 TGATCCCACTTACCTGAGCAAGAAACAGAAAGCCCAAGGAGTCCACAGTGTCT 1285  
Db 1033 TGATCCCACTTACCTGAGCAAGAAACAGAAAGCCCAAGGAGTCCACAGTGTCT 1092  
QY 1286 ATGAAGACCAACCACTGCAAGGCTGCGTGAATGTTGATGGGAAAGATTTTCAAG 1345  
Db 1093 ATGAAGACCAACCACTGCAAGGCTGCGTGAATGTTGATGGGAAAGATTTTCAAG 1152  
QY 1346 GGAAGCAACTTAAAGTCTCCCTTGTGCGAAGAAAGCTCCATATGAACAGTATGCGGGTGT 1405  
Db 1153 GGAAGCAACTTAAAGTCTCCCTTGTGCGAAGAAAGCTCCATATGAACAGTATGCGGGTGT 1212  
QY 1406 GTTCCCAACCCCGTGAAGGAGAGAGCAATGCCACCACTCCGTGAGTCCAGAGAGCC 1465  
Db 1213 GTTCCCAACCCCGTGAAGGAGAGAGCAATGCCACCACTCCGTGAGTCCAGAGAGCC 1272  
QY 1466 CAGAGAGTCTGAGGAGAGCCCATGAGTGCATGAGAGGCGGAGAGAGATGAGAGAGCT 1525  
Db 1273 CAGAGAGTCTGAGGAGAGCCCATGAGTGCATGAGAGGCGGAGAGAGATGAGAGAGCT 1332  
QY 1526 TCCCTCCAAAGAGAGCCCGGGGTTCGAGAGGAAACCCCTCTGAGAGAGAGAACTCCAGC 1585  
Db 1333 TCCCTCCAAAGAGAGCCCGGGGTTCGAGAGGAAACCCCTCTGAGAGAGAGAACTCCAGC 1392

QY 1586 ACCGAGCTGAGAGACTGAGAGTGTCCCAATCCGGTTGTGAAACCAAGACTTCCCTGGA 1645  
Db 1393 ACCGAGCTGAGAGACTGAGAGTGTCCCAATCCGGTTGTGAAACCAAGACTTCCCTGGA 1452  
QY 1646 GAAACAGAGTCAACCAAGTGAAGGAGCCCAAGGCTTCTCCCGCAGACCTTTTC 1705  
Db 1453 GAAACAGAGTCAACCAAGTGAAGGAGCCCAAGGCTTCTCCCGCAGACCTTTTC 1512  
QY 1706 CGCCCGCGGAGTGTGATGCTGAGCAGAGTGTGCTGAGTGCATGCGGAGAGAAAGAGGTG 1765  
Db 1513 CGCCCGCGGAGTGTGATGCTGAGCAGAGTGTGCTGAGTGCATGCGGAGAGAAAGAGGTG 1572  
QY 1766 GCTCATGAGATGCGTGTGTGTCCCGTGTGAAATGTTCAAGAGTGTGCGTGTGTGAGACAGAG 1825  
Db 1573 GCTCATGAGATGCGTGTGTGTCCCGTGTGAAATGTTCAAGAGTGTGCGTGTGTGAGACAGAG 1632  
QY 1826 GTGAGCTTCCGTTGTGTGCGGAGGAGCAGAGAGTGTGCTTGTGTGAGAGAGAGAGAGTGTG 1885  
Db 1633 GTGAGCTTCCGTTGTGTGCGGAGGAGCAGAGAGTGTGCTTGTGTGAGAGAGAGAGAGTGTG 1692  
QY 1886 GCCCTGGAGGAGCCCTTGAACCTTTGATGAAACAGATGAGAGAGAGAGAGAGAGTGTG 1945  
Db 1693 GCCCTGGAGGAGCCCTTGAACCTTTGATGAAACAGATGAGAGAGAGAGAGAGAGAGTGTG 1752  
QY 1946 GAGGACCTGGAAGAAATGATTAAGCGAGCAGCCTGACAGAGCGGAGAGATCGGCCCTACT 2005  
Db 1753 GAGGACCTGGAAGAAATGATTAAGCGAGCAGCCTGACAGAGCGGAGAGATCGGCCCTACT 1812  
QY 2006 AGATGACAGAGACCCCGAGCA 2025  
Db 1813 AGATGACAGAGACCCCGAGCA 1832

## RESULT 14

ADRO7446  
ID ADRO7446 standard; cDNA; 2026 BP.

AC ADRO7446;

DT 04-NOV-2004 (first entry)

DE Full length human cDNA useful for treating neurological disease Seq 952.

XX gene; ss; human; oligo-capping method; diagnostic marker; gene therapy;

KW osteoporosis; neurological disease; Alzheimer's disease;

KW Parkinson's disease; dementia; short memory; cancer;

KW sense or motor function; emotional reaction; fear response; panic;

KW osteopathic; neuroprotective; nootropic; antiparkinsonian; cyostatic;

XX tranquiliser.

OS Homo sapiens.

OS Homo sapiens.

PN EPI447413-A2.

PD 18-AUG-2004.

XX 12-FEB-2004; 2004EP-00003145.

PF 14-FEB-2003; 2003JP-00102207.

PR 09-MAY-2003; 2003JP-00131452.

XX (REAS-) RES ASSOC BIOTECHNOLOGY.

PA Isogai T, Yamamoto J, Nishikawa T, Isono Y, Sugiyama T, Otsuki T,

PI Makamatsu A, Ishii S, Nagai K, Irie R;

DR WPI; 2004-583265/57.

DR P-PSDB; ADRO9402.

XX New 1995 cDNA, useful for treating osteoporosis, neurological diseases,

PT Alzheimer's diseases, Parkinson's diseases, dementia and various cancers.

XX Claim 1; SEQ ID NO 952; 2686bp; English.



Oy	1801	GGAGGTGGCCCGTGGTGGAGACAGAGGTGACCTCCGTGGTGGCCCGGGGACATGACCCGAGGT	1860
Db	1646	AGAGGTGCCCTGTGTGGAGACAGAGGTGACTCTCCGTGGTGGCCCGGGGATGACCCGAGGT	1705
Oy	1861	GGCTTTGGTGGAGGAGAGCGAGGTGGCCCTCGGGGGGCCCTCGAGCCTTTGATGGACAG	1920
Db	1706	GGCTTTGGTGGAGGAGAGCGAGGTGGCCCTGGGGGGGCCCTCGAGCCTTTATGGACAG	1765
Oy	1921	ATGGGAGGAGAGAGAGAGACGTGGAGGACCTGGAAAAATGATTAAGGCGAGCACCGT	1980
Db	1766	ATGGGAGGAGAGAGAGAGACGTGGAGGACCTGGAAAAATGATTAAGGCGAGCACCGT	1825
Oy	1981	CAGGAGCCAGAGATCGGCCCTACTAGATGGAGAGACCCCGCAG	2025
Db	1826	CAGAGCCAGAGATCGGCCCTACTAGATGGAGAGACCCCGCAG	1870

RESULT 15  
 ADP56333  
 ID ADP56333 standard; cDNA; 1807 BP.  
 XX  
 AC ADP56333;  
 XX  
 DT 18-NOV-2004 (first entry)  
 XX  
 DE Human PRO cDNA sequence SEQ ID NO:2309.  
 XX  
 KM human; PRO; immune related disease; inflammatory immune response;  
 KM immune response stimulation; antiallergic; antianaemic; antiarthritic;  
 KM antidiabetic; antidiabetic; antiinflammatory; antipsoriatic;  
 KM antirheumatic; antihypoid; CNS; dermatological; gastrointestinal;  
 KM haemostatic; hepatotropic; immunostimulant; immunosuppressive; muscular;  
 KM neurotropic; neuroprotective; osteopathic; respiratory; vasotropic;  
 KM vitruide; gene therapy; gene; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200403956-A2.  
 XX  
 PD 13-MAY-2004.  
 XX  
 PF 28-OCT-2003; 2003WO-US034381.  
 XX  
 PR 29-OCT-2002; 2002US-0422472P.  
 XX  
 PA (GETH ) GENENTECH INC.  
 XX  
 PI Aggarwal S, Clark H, Gurney AL, Schoenfeld J, Williams PM;  
 PI Wood WI, Wu TD;  
 XX  
 DR WPI; 2004-376182/35.  
 DR P-PSDB; ADP56334.  
 PT  
 PT New PRO polynucleotides and polypeptides, useful in useful in diagnosing  
 PT and treating an immune related disease, e.g. systemic lupus  
 PT erythematosus, rheumatoid arthritis, diabetes mellitus or asthma and in  
 PT stimulating an immune response.  
 XX  
 Claim 2; SEQ ID NO 2309; 3009bp; English.  
 XX  
 XX The present invention describes an isolated PRO nucleic acid (I). Also  
 CC described: (1) a vector comprising (II); (2) a host cell comprising the  
 CC vector of (1); (3) a process for producing a PRO polypeptide; (4) an  
 CC isolated PRO polypeptide; (5) a chimeric molecule comprising the  
 CC polypeptide of (4) fused to a heterologous amino acid sequence; (6) an  
 CC antibody which specifically binds to a polypeptide of (4); (7) a  
 CC composition of matter comprising a polypeptide of (4), an agonist or  
 CC antagonist of the polypeptide or an antibody that binds to the  
 CC polypeptide in combination with a carrier; (8) an article of manufacture  
 CC comprising a container, a label on the container and a composition of  
 CC matter of (7); (9) a method of treating an immune related disease in a  
 CC mammal; (10) a method for determining the presence of a PRO polypeptide  
 CC in a sample suspected of having the polypeptide; (11) a method of

CC diagnosing an immune related disease or an inflammatory immune response  
CC in mammal; (12) a method of identifying a compound that inhibits or  
CC mimics the activity of or expression of a gene encoding a PRO polypeptide  
CC ; and (13) a method of stimulating the immune response in a mammal. The  
CC PRO sequences have antiallergic, antianemic, antarthritic,  
CC antiasthmatic, antidiabetic, antiinflammatory, antipruritic,  
CC antirheumatic, antichryoid, CNS, dermatological, gastrointestinal,  
CC haemostatic, hepatotropic, immunostimulant, immunosuppressive, muscular,  
CC nephrotropic, neuroprotective, osteopathic, respiratory, vasotonic and  
CC vincidic activities, and can be used in gene therapy. The nucleic acid  
CC (I) and the encoded polypeptides, compositions, kits and methods are  
CC useful in diagnosing and treating an immune related disease and in  
CC stimulating an immune response. The present sequence represents a human  
CC PRO nucleotide sequence from the present invention.  
XX  
SQ Sequence 1807 BP; 476 A; 497 C; 523 G; 311 T; 0 U; 0 Other;

Query Match	76.9%;	Score 1557.4;	DB 13;	Length 1807;
Best Local Similarity	89.0%;	Pred. No. 0;		
Matches 1787;	Conservative	0;	Mismatches 1;	Indels 219;
			Gaps 1;	

OY	1	GACGACGTTAGAGAAACGAGAGAAAGAGAAAATGGGCTCCACGGATTACAGTACC	60
Db	20	GACGACGTTAGAGAAACGAGAGAAAGAGAAAATGGGCTCCACGGATTACAGTACC	79
OY	61	TATAGCCAGCTGCAGCGCAGACGAGGCTTACAGTCTTACACCGGCCAGGCCACTCAAGA	120
Db	80	TATAGCCAGCTGCAGCGCAGACGAGGCTTACAGTCTTACACCGGCCAGGCCACTCAAGA	139
OY	121	TATGCAACAGACCCCAAGCATATGGGCAACAAAGCTATGGAACTTATGACAGCCCACT	180
Db	140	TATGCAACAGACCCCAAGCATATGGGCAACAAAGCTATGGAACTTATGACAGCCCACT	199
OY	181	GATGTCAGCTATACCCAGGCTCAGACCACTGACAACTATGGGCGACACCGCTTATGCACT	240
Db	200	GATGTCAGCTATACCCAGGCTCAGACCACTGACAACTATGGGCGACACCGCTTATGCACT	259
OY	241	TCTTATGAGACGCGCTCCCACTGGTTATATCTACTCAACTGCCCCCAGGCAATACAGCCAG	300
Db	260	TCTTATGAGACGCGCTCCCACTGGTTATATCTACTCAACTGCCCCCAGGCAATACAGCCAG	319
OY	301	CCTGTCCAGGGGTTATGGGCACTGTGTGCTTATGATACCAACACTGCTACAGTACACCAACC	360
Db	320	CCTGTCCAGGGGTTATGGGCACTGTGTGCTTATGATACCAACACTGCTACAGTACACCAACC	379
OY	361	CAGGCGTCTATGAGAGCTCAGTGTGCATATGAGCACTAGCGCTGCTTATCCAGCCTATGGG	420
Db	380	CAGGCGTCTATGAGAGCTCAGTGTGCATATGAGCACTAGCGCTGCTTATCCAGCCTATGGG	439
OY	421	CAGCAGCCAGACGCACTGCACTTACAAAGACCGAGAGTGGAAAACAAGCCCACTGAGACT	480
Db	440	CAGCAGCCAGACGCACTGCACTTACAAAGACCGAGAGTGGAAAACAAGCCCACTGAGACT	499
OY	481	AGTCAACTCATCTAGACACAGGGGGTTTACAACAGGCCAGCTTAGATATGACAGAGT	540
Db	500	AGTCAACTCATCTAGACACAGGGGGTTTACAACAGGCCAGCTTAGATATGACAGAGT	559
OY	541	AAGTCAAGTTATCCCAAGGTACTGGGAGCTAACCCCACTGACAGCCAGTCACTGCACCTCA	600
Db	560	AAGTCAAGTTATCCCAAGGTACTGGGAGCTAACCCCACTGACAGCCAGTCACTGCACCTCA	619
OY	601	TGCTAACCTCTCTACAGCTATTCTCTTACAACAGCCGACTAGTTATGATCAGAGCAGTTAC	660
Db	620	TGCTAACCTCTCTACAGCTATTCTCTTACAACAGCCGACTAGTTATGATCAGAGCAGTTAC	679
OY	661	TGTCAGCAGAACACCTTATGGGCAACCGAGAGCTATGGAACAGCAGAGTAAAGTATGGTCAA	720
Db	680	TGTCAGCAGAACACCTTATGGGCAACCGAGAGCTATGGAACAGCAGAGTAAAGTATGGTCAA	739
OY	721	CAAGCAGAGCTATGGGACAGAGCTTCCACATGATATACCAACCCCAAACTGATATCTTACAGC	780
Db	740	CAAGCAGAGCTATGGGACAGAGCTTCCACATGATATACCAACCCCAAACTGATATCTTACAGC	799

QY 781 CAAGCTCCAGTCAATATAGCAACAGAGCAGCTACGCGGACAGATTCATTCGCA 840  
DB 800 CAAGCTCCAGTCAATATAGCAACAGAGCAGCTACGCGGACAGCA----- 848  
QY 841 CAGAGCAACCCCAAGTCAATGAGTCTGAGGAGAGTCTGAGAGATTTTCCGACCA 900  
DB 849 ----- 848  
QY 901 GAGAGAACCCGAGCATGATGAGCTTGAATACCGGAGGACAGGAGGATTTGAT 960  
DB 849 ----- 848  
QY 961 CGTGAAGCAYTAGACAGAGTGGCGGAGAGAGACGCGTGAATGGCAGCGCTGA 1020  
DB 849 ----- 848  
QY 1021 GAGCAGAGTGGCTTCAATAAGCCTGTGAGCCCATGATGAAGACAGATCTTGATCTA 1080  
DB 849 -----GACCATGATGAAGACAGATCTTGATCTA 880  
QY 1081 GGGCCACCTGTAGATCCAGATGAAGATCTGACAAAGTGCATTTATGTCAAGATTA 1140  
DB 881 GGGCCCTCTGTAGATCCAGATGAAGATCTGACAAAGTGCATTTATGTCAAGATTA 940  
QY 1141 AATGACAGTGAATCTAGATGAATCTGACAACTTCTTAAGCAGTGGGAGTTGTTAAG 1200  
DB 941 AATGACAGTGAATCTAGATGAATCTGACAACTTCTTAAGCAGTGGGAGTTGTTAAG 1000  
QY 1201 ATGAAACAAGAACTGGGCAACCCATGATCCATCTACCTGACAAAGAAACAGAAAG 1260  
DB 1001 ATGAAACAAGAACTGGGCAACCCATGATCCATCTACCTGACAAAGAAACAGAAAG 1060  
QY 1261 CCCAAAGGCGATGCAAGTGTCTTAAGAAACCCACCTGCAAGGCTGCGGTGAA 1320  
DB 1061 CCCAAAGGCGATGCAAGTGTCTTAAGAAACCCACCTGCAAGGCTGCGGTGAA 1120  
QY 1321 TGGTTTATGGGAAAGATTTTCAAGGAGCAAACTTAAGTCTCTTGTCTCGAAGAG 1380  
DB 1121 TGGTTTATGGGAAAGATTTTCAAGGAGCAAACTTAAGTCTCTTGTCTCGAAGAG 1180  
QY 1381 CTTCAATGAACAGTATGCGGGGTGTCTGCAACCCGTAAGGAGGCAATGCCACA 1440  
DB 1181 CTTCAATGAACAGTATGCGGGGTGTCTGCAACCCGTAAGGAGGCAATGCCACA 1240  
QY 1441 CCACTCCGTGAGAGTCCAGAGGCGCCAGAGGCTCTGGGGGACCCATGGTCCGATGGGA 1500  
DB 1241 CCACTCCGTGAGAGTCCAGAGGCGCCAGAGGCTCTGGGGGACCCATGGTCCGATGGGA 1300  
QY 1501 GGCCTGAGAGAGATAGAGAGGCTTCTCTCAAGAGGACCCCGGGGTCCCGAGGGAGC 1560  
DB 1301 GGCCTGAGAGAGATAGAGAGGCTTCTCTCAAGAGGACCCCGGGGTCCCGAGGGAGC 1360  
QY 1561 CCTCTGAGAGAGAAAGTCCAGACCGAGCTGAGACTGAGAGTGTCCCAATCCGGGT 1620  
DB 1361 CCTCTGAGAGAGAAAGTCCAGACCGAGCTGAGACTGAGAGTGTCCCAATCCGGGT 1420  
QY 1621 TGTGGAACCAAGAACTTCCCTGAGAAACAGAGTGAACCAAGTGAAGGCCCAAGCT 1680  
DB 1421 TGTGGAACCAAGAACTTCCCTGAGAAACAGAGTGAACCAAGTGAAGGCCCAAGCT 1480  
QY 1681 GAAAGGCTTCTCCCGCACCTTCCGCGCCCGGGGTGTGATCGTGGCAGAGGTGGCCT 1740  
DB 1481 GAAAGGCTTCTCCCGCACCTTCCGCGCCCGGGGTGTGATCGTGGCAGAGGTGGCCT 1540  
QY 1741 GGTGGCATGCGGGAGAGAAAGTGGCTCAATGATCGTGGTGTCCCGGTGAATGTTT 1800  
DB 1541 GGTGGCATGCGGGAGAGAAAGTGGCTCAATGATCGTGGTGTCCCGGTGAATGTTT 1600  
QY 1801 AGAGGTGGCGGTGTGAGACAGAGTGGCTTCCGTGTGGCCGAGGCAATGAACCGAGGT 1860  
DB 1601 AGAGGTGGCGGTGTGAGACAGAGTGGCTTCCGTGTGGCCGAGGCAATGAACCGAGGT 1660  
QY 1861 GGCCTTGTGAGAGAGACAGAGTGGCCCTGGGGGGCCCTGTGACCTTTGATGAACAG 1920

DB 1661 GGCCTTGTGAGAGAGACAGAGTGGCCCTGGGGGGCCCTGTGACCTTTGATGAACAG 1720  
QY 1921 ATGGAGGAAAGAAAGAGAGAGAGCTGAGAGACTTGAAATAAGGCAAGCACTGT 1980  
DB 1721 ATGGAGGAAAGAAAGAGAGAGAGCTGAGAGAGCTTGAAATAAGGCAAGCACTGT 1780  
QY 1981 CAGAGCGCAGAGATGCGCCCTTACTAG 2007  
DB 1781 CAGAGCGCAGAGATGCGCCCTTACTAG 1807

Search completed: February 20, 2005, 05:30:45  
Job time : 997.09 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 20, 2005, 04:13:53; Search time 335.821 Seconds  
(without alignments)  
9866.749 Million cell updates/sec

Title: US-10-791-017a-1\_COPY\_8\_2032

Perfect score: 2025  
Sequence: 1 gaacgacgttgagagacga.....agatgcagagacccgcagaga 2025

Scoring table: IDENTITY\_NUC  
Gapop 10.0, Gapext 1.0

Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database: Issued Patents NA: \*  
1: /cgn2\_6/ptodata/1/ina/5A.COMB.seq: \*  
2: /cgn2\_6/ptodata/1/ina/5B.COMB.seq: \*  
3: /cgn2\_6/ptodata/1/ina/6A.COMB.seq: \*  
4: /cgn2\_6/ptodata/1/ina/6B.COMB.seq: \*  
5: /cgn2\_6/ptodata/1/ina/PTCUS.COMB.seq: \*  
6: /cgn2\_6/ptodata/1/ina/backfile1.seq: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2011.4	99.3	2371	2 US-08-343-443B-1	Sequence 1, Appli
2	1884.4	93.1	6002	4 US-09-949-016-13696	Sequence 13696, A
3	1556	76.8	1785	4 US-09-949-016-5043	Sequence 5043, Ap
4	1556	76.8	1785	4 US-09-949-016-5044	Sequence 5044, Ap
5	1429.6	70.6	1783	4 US-09-949-016-1954	Sequence 1954, Ap
6	819	40.4	2412	1 US-08-437-027-18	Sequence 18, Appl
7	424	20.9	454	4 US-09-513-999C-1657	Sequence 1657, Ap
8	364	18.0	601	4 US-09-949-016-66382	Sequence 66382, A
9	338	16.7	411	4 US-08-621-876-13361	Sequence 13361, A
10	299.2	14.8	393	4 US-09-513-999C-12062	Sequence 12062, A
11	266.4	13.2	954	2 US-08-343-443B-106	Sequence 106, App
12	258.4	12.8	35784	4 US-09-949-016-16785	Sequence 16785, A
13	258.4	12.8	35784	4 US-09-949-016-16786	Sequence 16786, A
14	213	10.5	601	4 US-09-949-016-176641	Sequence 176641,
15	213	10.5	601	4 US-09-949-016-176644	Sequence 176644,
16	173.2	8.7	1939	4 US-09-919-039-322	Sequence 322, App
17	135.8	6.7	601	4 US-09-949-016-176637	Sequence 176637,
18	135.8	6.7	601	4 US-09-949-016-176638	Sequence 176638,
19	135.8	6.7	601	4 US-09-949-016-176660	Sequence 176660,
20	135.8	6.7	601	4 US-09-949-016-176661	Sequence 176661,
21	129.4	6.4	601	4 US-09-949-016-176622	Sequence 176622,
22	129.4	6.4	601	4 US-09-949-016-176623	Sequence 176623,
23	129.4	6.4	601	4 US-09-949-016-176645	Sequence 176645,
24	129.4	6.4	601	4 US-09-949-016-176646	Sequence 176646,
25	116.4	5.7	450	3 US-09-370-838-145	Sequence 145, App
26	116.4	5.7	450	4 US-09-854-133-145	Sequence 145, App
27	74	3.7	114793	4 US-10-148-806-3	Sequence 3, Appli

28	70.2	3.5	215	4 US-09-513-999C-31510	Sequence 31510, A
29	67	3.3	152132	4 US-09-949-016-13885	Sequence 13885, A
30	67	3.3	152145	4 US-09-949-016-12371	Sequence 12371, A
31	66.4	3.3	311	4 US-09-313-294A-5928	Sequence 5928, Ap
32	65.6	3.2	7218	1 US-08-232-463-14	Sequence 14, Appl
33	64.6	3.2	601	4 US-09-949-016-176640	Sequence 176640,
34	64.6	3.2	601	4 US-09-949-016-176663	Sequence 176663,
35	60	3.0	268	1 US-08-437-027-19	Sequence 19, Appl
36	59.8	3.0	1505	1 US-07-915-246-1	Sequence 1, Appl
37	59	2.9	1225	4 US-09-976-594-416	Sequence 416, App
38	58.8	2.9	34230	4 US-09-949-016-12052	Sequence 12052, A
39	58.8	2.9	128470	4 US-09-949-016-13765	Sequence 13765, A
40	58.4	2.9	425	2 US-08-343-443B-45	Sequence 45, Appl
41	58	2.9	328	2 US-08-343-443B-5	Sequence 5, Appl
42	57.8	2.9	12695	4 US-09-949-016-16775	Sequence 16775, A
43	57.2	2.8	1682	4 US-09-220-132-82	Sequence 82, Appl
44	57.2	2.8	1684	4 US-09-919-039-323	Sequence 323, App
45	57	2.8	1926	3 US-09-249-585A-4	Sequence 4, Appli

#### ALIGNMENTS

RESULT 1  
US-08-343-443B-1  
Sequence 1, Application US/08343443B  
Patent No. 5968734  
GENERAL INFORMATION:  
APPLICANT: Aurias, Alain  
APPLICANT: Delattre, Olivier  
APPLICANT: Desmarte, Chantal  
APPLICANT: Melot, Thomas  
APPLICANT: Peter, Martine  
APPLICANT: Ploougastel, Beatrice  
APPLICANT: Thomas, Gilles  
APPLICANT: Zucman, Jessica  
TITLE OF INVENTION: NUCLEIC ACID CORRESPONDING TO A GENE OF  
TITLE OF INVENTION: CHROMOSOME 22 INVOLVED IN RECURRENT CHROMOSOMAL  
TITLE OF INVENTION: TRANSLATIONS ASSOCIATED WITH THE DEVELOPMENT OF CANCEROUS  
TITLE OF INVENTION: TUMORS, AND NUCLEIC ACIDS OF FUSION RESULTING FROM SAID  
TITLE OF INVENTION: TRANSLOCATIONS  
NUMBER OF SEQUENCES: 129  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Weiser & Associates  
STREET: 230 South Fifteenth Street  
CITY: Philadelphia  
STATE: PA  
COUNTRY: USA  
ZIP: 19102  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: AEDIT 1.0 DOS text editor  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/343,443B  
FILING DATE: 18-NOV-1994  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/FR93/00494  
FILING DATE: 19-MAY-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: FR 92/06123  
FILING DATE: 20-MAY-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Weiser, Gerard J.  
REGISTRATION NUMBER: 19,763  
REFERENCE/DOCKET NUMBER: 989,6121P  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 215-875-8383  
TELEFAX: 215-875-8394  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:

```

; LENGTH: 2371 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
;
FEATURE:
; NAME/KEY: CDS
; LOCATION: 25..1992
;
US-08-343-443B-1

```

Query Match	99.3%;	Score 2011.4;	DB 2;	length 2371;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 2012; Conservative	0;	Mismatches 1;	Indels 0;	Gaps 0;

OY	13	GAGAACGAGGAGGAGGAGGAGAAATGGGCTCAGCGATTACGTACCTATAGCCAAAGT	72
Db	1	GAGAACGAGGAGGAGGAGGAGAAATGGGCTCAGCGATTACGTACCTATAGCCAAAGT	60
OY	73	GCAGGCGAGCAGGAGGCTACAGTCTTACACCGCCAGCCCACTCAAGATATGACAGAAC	132
Db	61	GCAGGCGAGCAGGAGGCTACAGTCTTACACCGCCAGCCCACTCAAGATATGACAGAAC	120
OY	133	ACCCAGGCAATATGGGCAACAAAGCTATGGAACCTATGGAACGCCCACTGATGCGATAT	192
Db	121	ACCCAGGCAATATGGGCAACAAAGCTATGGAACCTATGGAACGCCCACTGATGCGATAT	180
OY	193	ACCCAGGCTCAGACCACTGCACACTATGGGCGAGACCGCCTATGCAACTTCTTATGACAG	252
Db	181	ACCCAGGCTCAGACCACTGCACACTATGGGCGAGACCGCCTATGCAACTTCTTATGACAG	240
OY	253	CCTCCCACTGGTTATATCTACTCCAACTGCCCCCCAGGCAATACGCCAGCTGTCCAGGGG	312
Db	241	CCTCCCACTGGTTATATCTACTCCAACTGCCCCCCAGGCAATACGCCAGCTGTCCAGGGG	300
OY	313	TATGSCACTGGGCTTATGATACCAACACTGCTACAGTCAACCAACCCAGGCTCTTAT	372
Db	301	TATGSCACTGGGCTTATGATACCAACACTGCTACAGTCAACCAACCCAGGCTCTTAT	360
OY	373	GCAGCTCAGTCTGCATATGSCACTCAGCCTGCTTATCCAGCCTATGSGCAGACGCCAGCA	432
Db	361	GCAGCTCAGTCTGCATATGSCACTCAGCCTGCTTATCCAGCCTATGSGCAGACGCCAGCA	420
OY	433	GCCATGCAACCTACAAAGACCGGAGATGGGAAACAAGCCCACTGAGACTTATGCAACCTCAA	492
Db	421	GCCATGCAACCTACAAAGACCGGAGATGGGAAACAAGCCCACTGAGACTTATGCAACCTCAA	480
OY	493	TCTAGCACAGGGGGTTACAAACCAAGCCCAAGCTATGATACAGATACAGATACCTAAGTTAT	552
Db	481	TCTAGCACAGGGGGTTACAAACCAAGCCCAAGCTATGATACAGATACAGATACCTAAGTTAT	540
OY	553	CCCCAGGTACCTTGGGAGCTAACCCATNGACGCCAGTCACTGCACCTCCATCCCTCACTCCT	612
Db	541	CCCCAGGTACCTTGGGAGCTAACCCATNGACGCCAGTCACTGCACCTCCATCCCTCACTCCT	600
OY	613	ACCAGCTATTCCTCTACACAGCCGACTAGTTATGATCAGAGCAGTTACTCTCAGCAGAAC	672
Db	601	ACCAGCTATTCCTCTACACAGCCGACTAGTTATGATCAGAGCAGTTACTCTCAGCAGAAC	660
OY	673	ACCTATGGGCAACCGAGCAGCTATGSCACGACAGATAGCTATGTTAAACAAAGCAGCTAT	732
Db	661	ACCTATGGGCAACCGAGCAGCTATGSCACGACAGATAGCTATGTTAAACAAAGCAGCTAT	720
OY	733	GGGCGAGCAGCTCCCACTAGTTATACCAACCCCAAACCTGATNCTTACAGCCCAAGTCCAAAT	792
Db	721	GGGCGAGCAGCTCCCACTAGTTATACCAACCCCAAACCTGATNCTTACAGCCCAAGTCCAAAT	780
OY	793	CAATATAGCCAAACAGAGCAGACGCTACGGGCGACAGAGTTCAATTCGACAGGACCAACCCC	852
Db	781	CAATATAGCCAAACAGAGCAGACGCTACGGGCGACAGAGTTCAATTCGACAGGACCAACCCC	840
OY	853	AGTACATACGGGTGTTTATATGGGCGAGAGTCTGAGAGATTTTCCGACACAGAGAGAACCCGG	912
Db	841	AGTACATACGGGTGTTTATATGGGCGAGAGTCTGAGAGATTTTCCGACACAGAGAGAACCCGG	900

OY	913	AGCATGATGTGGCCCTGATTAAACCGGGGACGGGGAAGAAGGGGGATTTCATCTGTGAGGCATG	972
Dd	901	AGCATTGATGTGCCCTTGATTAAACCGGGGACGGGGAAGAAGGGGGATTTCATCTGTGAGGCATG	960
OY	973	AGCAGAGTGGGCGGGGAGGAGGACCGCGGTGGAAATGGGACAGCGCTGGAGACGAGGTGGC	1032
Dd	961	AGCAGAGTGGGCGGGGAGGAGGACCGCGGTGGAAATGGGACAGCGCTGGAGACGAGGTGGC	1020
OY	1033	TTCAATAAAGCTTGGTGGACCCTATGGAATGMAAGAACCAATCTTGAATCTTAAGCCCCACTGTA	1092
Dd	1021	TTCAATAAAGCTTGGTGGACCCTATGGAATGMAAGAACCAATCTTGAATCTTAAGCCCCACTGTA	1080
OY	1093	GATCCAGATGNAAGACTCTGACAACACAGTGCATATTATGTACAAGGATTAAATGACAGTGTG	1152
Dd	1081	GATCCAGATGNAAGACTCTGACAACACAGTGCATATTATGTACAAGGATTAAATGACAGTGTG	1140
OY	1153	ACTTAGATGATCTGGCAGACTTCTTTTAAAGCAGTGGGGGTGTGTAAGATGAACAAGAGA	1212
Dd	1141	ACTTAGATGATCTGGCAGACTTCTTTTAAAGCAGTGGGGGTGTGTAAGATGAACAAGAGA	1200
OY	1213	ACTGGGCAACCCATGATTCACATTTACTTGGACAAGAAACAGAAAGCCCMAAGCGGAT	1272
Dd	1201	ACTGGGCAACCCATGATTCACATTTACTTGGACAAGAAACAGAAAGCCCMAAGCGGAT	1260
OY	1273	GCCACAGTCTCTATGAAAGCCCAACCACTGCCAAGGCTGCGGTGAATGTTGATGGG	1332
Dd	1261	GCCACAGTCTCTATGAAAGCCCAACCACTGCCAAGGCTGCGGTGAATGTTGATGGG	1320
OY	1333	AAGAATTTTCAMGGGAGCAAACTTAAAGTCTCCCTTGCTCGGAAGAAAGCTCCATGATAC	1392
Dd	1321	AAGAATTTTCAMGGGAGCAAACTTAAAGTCTCCCTTGCTCGGAAGAAAGCTCCATGATAC	1380
OY	1393	AGTATGGGGGTGGTCTTGCCAACCCGTGAGGGCACAGGCATGCCACCACTCCGTGGA	1452
Dd	1381	AGTATGGGGGTGGTCTTGCCAACCCGTGAGGGCACAGGCATGCCACCACTCCGTGGA	1440
OY	1453	GSTCAAGAGGCCCCAGAGAGTCTGGGGGGAAACCAATGGGTGCAATGGGAGGCCGTGAGGA	1512
Dd	1441	GSTCAAGAGGCCCCAGAGAGTCTGGGGGGAAACCAATGGGTGCAATGGGAGGCCGTGAGGA	1500
OY	1513	GATAGAGAGGCTTCCCTCCMAAGAGGACCCCGGGGTTCCCGAGGGAAACCCCTCTGAGGA	1572
Dd	1501	GATAGAGAGGCTTCCCTCCMAAGAGGACCCCGGGGTTCCCGAGGGAAACCCCTCTGAGGA	1560
OY	1573	GGAACCGTCCAGCACCGAGCTGGAGACTGGCAGTGTCCCAATCCGGGTTGTGAAAACAG	1632
Dd	1561	GGAACCGTCCAGCACCGAGCTGGAGACTGGCAGTGTCCCAATCCGGGTTGTGAAAACAG	1620
OY	1633	AACTTGGCTGGAABAACAAGTGCACAACAGTAAAGGCCCCMAAGCCGMAAGGCTTCTC	1692
Dd	1621	AACTTGGCTGGAABAACAAGTGCACAACAGTAAAGGCCCCMAAGCCGMAAGGCTTCTC	1680
OY	1693	CCGCCACCTTCCGCCCCCGGGGTGTGATCTGTGGCAGAGGTGGCCCTGTGTGCATGGG	1752
Dd	1681	CCGCCACCTTCCGCCCCCGGGGTGTGATCTGTGGCAGAGGTGGCCCTGTGTGCATGGG	1740
OY	1753	GAAGGAAGGTGGCTTCATGATCTGTGTGTCTCCGGTGAATGTTCAAGGTGGCCGT	1812
Dd	1741	GAAGGAAGGTGGCTTCATGATCTGTGTGTCTCCGGTGAATGTTCAAGGTGGCCGT	1800
OY	1813	GGTGGAACAAGGTGGCTTCGTGTGTGGCCGGGGACATGACCCAGAGGTGGCTTGTGTGGA	1872
Dd	1801	GGTGGAACAAGGTGGCTTCGTGTGTGGCCGGGGACATGACCCAGAGGTGGCTTGTGTGGA	1860
OY	1873	GGAAGACAGAGTGGCCCTGGGGGGCCCCCTTGACCTTGTATGTAACAGATGGAGGAAGA	1932
Dd	1861	GGAAGACAGAGTGGCCCTGGGGGGCCCCCTTGACCTTGTATGTAACAGATGGAGGAAGA	1920
OY	1933	AGAGGAGACCTGAGGACCTTGAAAAATGATTAAGCGAGCACCGTCAAGAGCGCAGA	1992
Dd	1921	AGAGGAGACCTGAGGACCTTGAAAAATGATTAAGCGAGCACCGTCAAGAGCGCAGA	1980
OY	1993	GATGGCCCTTACTAAGATCAGAGACCCCCGACGA	2025



Db	3681	GGCTTCCCTCCCGCACCTTCCACCCCCCGGGGTGGTATCATGGCAGAGGTGGCCCTGGT	3740
QY	1744	GGCATGCCGGGAGGAGNAGGTGGCTCATGATCGTGGTGTCCCGGTGAATGTTCA	1803
Db	3741	GGCATGTGGGAGGAGAGGTGGCTCTCATGATCATGTGGTCCCGGTGAATGTTCA	3800
QY	1804	GGTGGCCGTGTGGAGACAGAGGTGGCTTCCGTGGTGGCCGGGACATGACCCGAGGTGGC	1863
Db	3801	GGTGGCTGTGGTAGACACAGAGGTGGCTTCCGTGGTGG-CTGGGCAATGACCGAGGTGGC	3855
QY	1864	TTTGGTGGAGACGACGAGTGGCCCTTGGGGGGCCCCCTTGACCTTTGATGAAACAGATG	1923
Db	3860	TTTGGTGGAGAAAGACAGGTGGCCCTGGGGGGCCCC-GGACCTTTGATGAAACAGATG	3918
QY	1924	GGAGGAAAGAGAGAGACGTGGAGAGACCTTGAAAAATGATTAAGCGAGACACCTGAG	1983
Db	3919	GGAGGAAAGAGAGAGACGTGAAGACCTTGAAAAAGATTAAGCGAGACACCTGAG	3978
QY	1984	GAGGCGAGAGATCGGCTCTACTAGATGCAGAGACCCGCGAGA	2025
Db	3979	GAGGCGAGAGATCAGCCCTTACTAGATGCAGAGACCCGCGAGA	4020

### RESULT 3

```

US-09-949-016-5043
Sequence 5043, Application US/09949016
Patent No. 6812339
GENERAL INFORMATION:
APPLICANT: VENTER, J. Craig et al.
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
FILE REFERENCE: CL001307
CURRENT APPLICATION NUMBER: US/09/949,016
CURRENT FILING DATE: 2000-04-14
PRIOR APPLICATION NUMBER: 60/241,755
PRIOR FILING DATE: 2000-10-20
PRIOR APPLICATION NUMBER: 60/237,768
PRIOR FILING DATE: 2000-10-03
PRIOR APPLICATION NUMBER: 60/231,498
PRIOR FILING DATE: 2000-09-08
NUMBER OF SEQ ID NOS: 207012
SOFTWARE: PasteSeq for Windows Version 4.0
SEQ ID NO 5043
LENGTH: 1785
TYPE: DNA
ORGANISM: Human
US-09-949-016-5043

```

Query Match	76.8%	Score 1556;	DB 4;	Length 1785;
Best Local Similarity	89.1%	Pred. No. 0;		
Matches 1785; Conservative	0;	Mismatches 0;	Indels 219;	Gaps 1;

QY	4	GGACGTTAGAGAACGAGAGGAAAGAAATGGGCTCCAGGATTAACGTACCTAT	63
Db	1	GGACGTTAGAGAACGAGAGGAAAGAAATGGGCTCCAGGATTAACGTACCTAT	60
QY	64	AGCCAAAGTCGACGGCCAGCAGGGCTTCAAGTCTTACACCGCCAGCCCACTCAAGATAT	123
Db	61	AGCCAAAGTCGACGGCCAGCAGGGCTTCAAGTCTTACACCGCCCACTCAAGATAT	120
QY	124	GCAAGACCAACCCAGGCAATATGGGCAACAAACCTATGGAACTATGSAACGGCCACTGAT	183
Db	121	GACAGACCAACCCAGGCAATATGGGCAACAAACCTATGGAACTATGSAACGGCCACTGAT	180
QY	184	GTCAAGCTATACCCAGGCTCAGACCACTGCACCTATGGGCAACCGCCTATGCAACTCT	243
Db	181	GTCAAGCTATACCCAGGCTCAGACCACTGCACCTATGGGCAACCGCCTATGCAACTCT	240
QY	244	TATGACACAGCTCCCACTAGTATTATCTACTCCACTGCCCCCAGGCATACAGCCAGCT	303
Db	241	TATGACACAGCTCCCACTAGTATTATCTACTCCACTGCCCCCAGGCATACAGCCAGCT	300

QY	304	GTCCAGGGGATATGGCACTGGTGGCTTAATGATACCAACCACTGTAAGTCAACCAACCCAG	363
Db	301	GTCCAGGGGATATGGCACTGGTGGCTTAATGATACCAACCACTGTAAGTCAACCAACCCAG	360
QY	364	GCCCTCTATGCAAGCTCAAGTCTGCATATGGAATCTCAAGCTTGCTTATCCAGCTTATGGCAG	423
Db	361	GCCCTCTATGCAAGCTCAAGTCTGCATATGGAATCTCAAGCTTGCTTATCCAGCTTATGGCAG	420
QY	424	CAGCCAGCAAGCACTGCACCTTACAAAGACCCAGAGATGGAAACAAAGCCCACTGAGACTAGT	483
Db	421	CAGCCAGCAAGCACTGCACCTTACAAAGACCCAGAGATGGAAACAAAGCCCACTGAGACTAGT	480
QY	484	CAACCTCAATCTAGCAAGGGGGTTAAACAGGCCAGCCAGCTAGAGATATGCAAGAGTAAC	543
Db	481	CAACCTCAATCTAGCAAGGGGGTTAAACAGGCCAGCTAGAGATATGCAAGAGTAAC	540
QY	544	TACAGTTATCCCAAGGTAACCTTGGAGGCTTACCCATGCAAGCCCATGCACTGCACCTCCATCC	603
Db	541	TACAGTTATCCCAAGGTAACCTTGGAGGCTTACCCATGCAAGCCCATGCACTGCACCTCCATCC	600
QY	604	TACCTCTCTACAGACTTATCTCTTACACAGCCGACTAGTTATGATTCAGACAGTAACTCT	663
Db	601	TACCTCTCTACAGACTTATCTCTTACACAGCCGACTAGTTATGATTCAGACAGTAACTCT	660
QY	664	CAGCAGAACACTTATGGGCAACCGAGAGCTATGGAACAGAGATAGCTTATGGCAACAA	723
Db	661	CAGCAGAACACTTATGGGCAACCGAGAGCTATGGAACAGAGATAGCTTATGGCAACAA	720
QY	724	AGCAGCTATGGGCGAGCAGCTCCCACTAGTTATCCCACTGGATCTTACAGCCAA	783
Db	721	AGCAGCTATGGGCGAGCAGCTCCCACTAGTTATCCCACTGGATCTTACAGCCAA	780
QY	784	GCTCCAGTCAATATAGCCAAACAGACAGCAAGCTTACGGGCGACGAGATTCACTCCGACAG	843
Db	781	GCTCCAGTCAATATAGCCAAACAGACAGCAAGCTTACGGGCGACGAGATTCACTCCGACAG	840
QY	844	GACCAACCCCAAGTACGATGGGTGTTATGGGAGAGATCTGGAGATTTTCCGACACAGA	903
Db	827	-----	826
QY	904	GAGAACCGGAGCATGATGGCCCTGATTAACCGGGGAGAGGAGATTTGATCGT	963
Db	827	-----	826
QY	964	GGAAGCATGACAGAGCTGGGCGGGGAGAGAGACCGGCTGGAATGGCGAGCCTGAGAG	1023
Db	827	-----	826
QY	1024	CGAGGTGCTTCAATTAAGCCGTGTGAGCCCATGGATGAAGGACAGATCTTGATCTTAGGC	1083
Db	827	-----GACCCATGGATTAAGGACCAAGATCTTGATCTTAGGC	861
QY	1084	CCACCTGATATCCAGATGAGAGCTCTGCACAAAGTGCATTTTATGTATCAAGATTAAAT	1143
Db	862	CCACCTGATATCCAGATGAGAGCTCTGCACAAAGTGCATTTTATGTATCAAGATTAAAT	921
QY	1144	GACAGTGTGACTTGAATGATCTGTGGCAGACTTCTTTAAGCAGTGTGGGTTGTTAAGATG	1203
Db	922	GACAGTGTGACTTGAATGATCTGTGGCAGACTTCTTTAAGCAGTGTGGGTTGTTAAGATG	981
QY	1204	AACAAAGAACTTGGGCAACCCATATCCACTTCACTGACCAAGAAACAGAAAGCC	1263
Db	982	AACAAAGAACTTGGGCAACCCATATCCACTTCACTGACCAAGAAACAGAAAGCC	1041
QY	1264	AAAGGCGATGCAAGATGCTTATGAAGACCCACCACTGCAAGGCTGCGGTGGAATGG	1323
Db	1042	AAAGGCGATGCAAGATGCTTATGAAGACCCACCACTGCAAGGCTGCGGTGGAATGG	1101
QY	1324	TTTGTATGGGAAGAATTTTCAAGGAGCAACTTAAAGTCTTCCCTTGTCTCGAAGAAAGCT	1383
Db	1102	TTTGTATGGGAAGAATTTTCAAGGAGCAACTTAAAGTCTTCCCTTGTCTCGAAGAAAGCT	1161
QY	1384	CCAAATGAACAGTATGCGGGGTGTCTGCAACCCCGTAGGGGCAAGGACATGCCACACCA	1443

Db	1162	CCAAATGAAACAATATGGGGGTGTCTGCCACCCCGTAGGGGACAGAGCATGCGACACCA	1221
QY	1444	CTCCGTGGAGGTTCCAGAGAGGCCCAAGAGGTCTTG9999AACCAT999TGCAT999A93C	1503
Db	1222	CTCCGTGGAGGTTCCAGAGAGGCCCAAGAGGTCTTG9999AACCAT999TGCAT999A93C	1281
QY	1504	CGTGAAGAGATAPAGAGAGGCTTCCCTCCAAAGAGACCCCGGGGTTTCCGAGGAAACCC	1565
Db	1282	CGTGAAGAGATAPAGAGAGGCTTCCCTCCAAAGAGACCCCGGGGTTTCCGAGGAAACCC	1341
QY	1564	TTTGGAGAGGAAACGTCCAGACCCGAGCTG9AGACTGGCAATGTCCCAATCCGGGTTGT	1622
Db	1342	TTTGGAGAGGAAACGTCCAGACCCGAGCTG9AGACTGGCAATGTCCCAATCCGGGTTGT	1401
QY	1624	GGAAACCAAGAACTTGGCCTCGAGAAACAGAGTCAACAGTGTAPAGGCCCCAAAGCTGAA	1683
Db	1402	GGAAACCAAGAACTTGGCCTCGAGAAACAGAGTCAACAGTGTAPAGGCCCCAAAGCTGAA	1461
QY	1684	GGCTTCCTCCCGCCACCCCTTTCCGCCCCCGGGTGTATCGTGGAGAGGTGGCCCTGTGT	1743
Db	1462	GGCTTCCTCCCGCCACCCCTTTCCGCCCCCGGGTGTATCGTGGAGAGGTGGCCCTGTGT	1521
QY	1744	GGCATGCGGGGAGGAAAGAGGTGACCTCATATGATCGTGGTGGTCCCGGTGGAATGTTCAGA	1801
Db	1522	GGCATGCGGGGAGGAAAGAGGTGACCTCATATGATCGTGGTGGTCCCGGTGGAATGTTCAGA	1581
QY	1804	GGTGGCCGTGTGTGAGACAGAGGTGGCTTCCGTGTGTGGCCG999CATG9ACCGAGTGGC	1863
Db	1582	GGTGGCCGTGTGTGAGACAGAGGTGGCTTCCGTGTGTGGCCG999CATG9ACCGAGTGGC	1641
QY	1864	TTTGGTGGAGGAAAGACGAGGTGGCCCTGGGGGGCCCCCTG9ACTTTTATATGAAACAGATG	1923
Db	1642	TTTGGTGGAGGAAAGACGAGGTGGCCCTGGGGGGCCCCCTG9ACTTTTATATGAAACAGATG	1701
QY	1924	GGAGGAAAGAGAGAGAGAGCTGTGAGAGACTGTGAAATAATGATTAAGCGAGACCGTTCAG	1983
Db	1702	GGAGGAAAGAGAGAGAGAGCTGTGAGAGACTGTGAAATAATGATTAAGCGAGACCGTTCAG	1761
QY	1984	GAGCGCAAGATCCGGCCCTACTACG 2007	
Db	1762	GAGCGCAAGATCCGGCCCTACTACG 1785	

```

RESULT 4
US-09-949-016-5044
: Sequence 5044, Application US/09949016
: Patent No. 6812339
: GENERAL INFORMATION:
: APPLICANT: VENTER, J. Craig et al.
: TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
: FILE REFERENCE: CLO01307
: CURRENT APPLICATION NUMBER: US/09/949,016
: CURRENT FILING DATE: 2000-04-14
: PRIOR APPLICATION NUMBER: 60/241,755
: PRIOR FILING DATE: 2000-10-20
: PRIOR APPLICATION NUMBER: 60/237,768
: PRIOR FILING DATE: 2000-10-03
: PRIOR APPLICATION NUMBER: 60/231,498
: PRIOR FILING DATE: 2000-09-08
: NUMBER OF SEQ ID NOS: 207012
: SOFTWARE: FASTSEQ for Windows Version 4.0
: SEQ ID NO 5044
: LENGTH: 1785
: TYPE: DNA
: ORGANISM: Human
US-09-949-016-5044

```

Query Match	76.8%	Score 1556;	DB 4;	Length 1785;
Best Local Similarity	89.1%	Pred. No. 0;		
Matches 1785; Conservative	0;	Mismatches 0;	Indels 219;	Gaps 14

OY	4	GGAGCTTGA	GAAACG	AGAGAGAA	AGAGAA	AAATGGG	GTCCAG	CGATTAC	AGTAC	CTAT	63
Db	1	GGACCTTGA	GAAACG	AGAGAGAA	AGAGAA	AAATGG	GTCCAG	CGATTAC	AGTAC	CTAT	60
OY	64	AGCCAGCTG	CAGCGC	CAGCAGG	CTACAGT	GGTCTAC	CCGCC	CAAGCCCA	CTCAAGATAT	123	
Db	61	AGCCAGCTG	CAGCGC	CACGGG	CTACAGT	GGTCTAC	CCGCC	CAAGCCCA	CTCAAGATAT	120	
OY	124	GCA	CAGACCA	CCCAAG	CAATATG	GGCA	CAAAAGCTAT	TGAA	CAAGCCCA	CTGAT	183
Db	121	GCA	CAGACCA	CCCAAG	CAATATG	GGCA	CAAAAGCTAT	TGAA	CAAGCCCA	CTGAT	180
OY	184	GT	CAGCTAT	CCCAAGG	CTCAGACCA	CTGCAAC	CTATATG	GGCAGACCG	CTATATG	CAACTTCT	243
Db	181	GT	CAGCTAT	CCCAAGG	CTCAGACCA	CTGCAAC	CTATATG	GGCAGACCG	CTATATG	CAACTTCT	240
OY	244	TAT	GACAGC	AGCTCCCA	CTGTTATAT	CTACT	CAACTG	CCCCCA	GGCATACAGCC	AGCT	303
Db	241	TAT	GACAGC	AGCTCCCA	CTGTTATAT	CTACT	CAACTG	CCCCCA	GGCATACAGCC	AGCT	300
OY	304	GT	CAAGGGG	GTATG	GCATG	TGTAT	GATATCA	CAACTG	CTTACATG	CAACCA	363
Db	301	GT	CAAGGGG	GTATG	GCATG	TGTAT	GATATCA	CAACTG	CTTACATG	CAACCA	360
OY	364	GC	CTCTTAT	TGACAG	CTCAGT	CTGCATAT	TGGCACT	CAGCCTG	CTTATCCAG	CCCTATATG	423
Db	361	GC	CTCTTAT	TGACAG	CTCAGT	CTGCATAT	TGGCACT	CAGCCTG	CTTATCCAG	CCCTATATG	420
OY	424	CAG	CCAGAGCA	CTGCA	CTTACAA	GACCG	AGATG	GGAA	CAAGCCCA	CTGAGCTAGT	483
Db	421	CAG	CCAGAGCA	CTGCA	CTTACAA	GACCG	AGATG	GGAA	CAAGCCCA	CTGAGCTAGT	480
OY	484	CA	ACCTCAAT	CTAGCA	CAGGGGG	TTTCA	ACCA	CAAGCCCA	CTAGATAT	GACAGATAC	543
Db	481	CA	ACCTCAAT	CTAGCA	CAGGGGG	TTTCA	ACCA	CAAGCCCA	CTAGATAT	GACAGATAC	540
OY	544	TAC	AGTTAT	CCCCAG	GTATC	CTGGG	AGCTAC	CCCCAT	GACGCCA	AGTAC	603
Db	541	TAC	AGTTAT	CCCCAG	GTATC	CTGGG	AGCTAC	CCCCAT	GACGCCA	AGTAC	600
OY	604	TAC	CCCTCTAC	CAGCTAT	TTCCCTCT	CA	CAGCGCA	CTAGTTAT	TGATCAG	AGAGGTTACT	663
Db	601	TAC	CCCTCTAC	CAGCTAT	TTCCCTCT	CA	CAGCGCA	CTAGTTAT	TGATCAG	AGAGGTTACT	660
OY	664	CAG	CAGAA	CA	ACTAT	TGGG	CAACCG	AGCCTAT	TGGA	CAGAGAT	723
Db	661	CAG	CAGAA	CA	ACTAT	TGGG	CAACCG	AGCCTAT	TGGA	CAGAGAT	720
OY	724	AG	CAGCTAT	TGGG	CAGC	AGCTCCCA	CTAT	AGTTACCA	CCCCAA	CTGATCT	783
Db	721	AG	CAGCTAT	TGGG	CAGC	AGCTCCCA	CTAT	AGTTACCA	CCCCAA	CTGATCT	780
OY	784	G	CTCAAGT	CAATAT	TAGCCAA	CAGAG	CA	CGGCA	CGTAC	CGGAGTT	843
Db	781	G	CTCAAGT	CAATAT	TAGCCAA	CAGAG	CA	CGGCA	CGTAC	CGGAGTT	826
OY	844	GAC	CA	CCCA	CTGAT	GCAT	TGGGT	TATATG	GGCAGAG	ATTTCC	903
Db	827	-----	-----	-----	-----	-----	-----	-----	-----	-----	826
OY	904	G	A	GAA	CCG	GAG	ATG	ATG	GGCC	TGAT	963
Db	827	-----	-----	-----	-----	-----	-----	-----	-----	-----	826
OY	964	G	A	GAG	CA	TG	A	GAG	GT	GGCGG	1023
Db	827	-----	-----	-----	-----	-----	-----	-----	-----	-----	826
OY	1024	C	G	A	G	A	G	T	G	A	1083
Db	827	-----	-----	-----	-----	-----	-----	-----	-----	-----	861
OY	1084	C	C	A	C	T	G	A	T	G	1143

```

Db      862  CCACCTGTAGATCAAGTGAAGACTCTGACAAAGTCAATTTATATGACAAAGATTAAAT 921
Qy      1144  GACAGTGTGACTCTAGATGATCTGGAGACTCTTTAAGAGTGTGGGTTGTTAAGATG 1203
Db      922  GACAGTGTGACTCTAGATGATCTGGAGACTCTTTAAGAGTGTGGGTTGTTAAGATG 981
Qy      1204  AACAAAGAACTGGGCAACCAATGATCTACCTGGAAGAAGAAAGAAAGCC 1263
Db      982  AACAAAGAACTGGGCAACCAATGATCTACCTGGAAGAAGAAAGAAAGCC 1041
Qy      1264  AAAAGCGATGCCACAGTGTCTTGAAGACCAACCACTGCAAGGCTGCGTGAATG 1323
Db      1042  AAAAGCGATGCCACAGTGTCTTGAAGACCAACCACTGCAAGGCTGCGTGAATG 1101
Qy      1324  TTTGATGGAAAGATTTTCAAGGGAGCAACTTAAATCTCTCTGCTCGAAGAACCT 1383
Db      1102  TTTGATGGAAAGATTTTCAAGGGAGCAACTTAAATCTCTCTGCTCGAAGAACCT 1161
Qy      1384  CCAATGAACAGTATGCGGGGTGCTGTCGCAACCCGTTAGGGGCAAGGCAATGCCA 1443
Db      1162  CCAATGAACAGTATGCGGGGTGCTGTCGCAACCCGTTAGGGGCAAGGCAATGCCA 1221
Qy      1444  CTCCTGGAAGTCCAGAGAGGCCAGAGATCTTGGGGGACCCATGAGTCCGATGGAGC 1503
Db      1222  CTCCTGGAAGTCCAGAGAGGCCAGAGATCTTGGGGGACCCATGAGTCCGATGGAGC 1281
Qy      1504  CGTGAAGAGATAGAGAGGCTTCTCCCAAGAGAACCCCGGGGTTCCCAAGGAAACCC 1563
Db      1282  CGTGAAGAGATAGAGAGGCTTCTCCCAAGAGAACCCCGGGGTTCCCAAGGAAACCC 1341
Qy      1564  TCTGAGAGAGAGAAAGTCCAGCACCGAGCTGGAAGTGGCAAGTCTCCAAATCCGGGTT 1623
Db      1342  TCTGAGAGAGAGAAAGTCCAGCACCGAGCTGGAAGTGGCAAGTCTCCAAATCCGGGTT 1401
Qy      1624  GGAACACGAGACTTCCGCTGGAAGACAGATGCAACAGTGAAGGCCCAAGACCTGAA 1683
Db      1402  GGAACACGAGACTTCCGCTGGAAGACAGATGCAACAGTGAAGGCCCAAGACCTGAA 1461
Qy      1684  GGCCTCTCCGCCACCCCTTCCGCCGCCCGGGTGTGATGTGGCAGAGGTGCGCTGT 1743
Db      1462  GGCCTCTCCGCCACCCCTTCCGCCGCCCGGGTGTGATGTGGCAGAGGTGCGCTGT 1521
Qy      1744  GGCATGCGGGAGAGAAAGGTGCTGATGATGATGATGATGATGATGATGATGATGATG 1803
Db      1522  GGCATGCGGGAGAGAAAGGTGCTGATGATGATGATGATGATGATGATGATGATGATG 1581
Qy      1804  GGTGGCCGTGTGAGACAGAGGTGCTTCCGTGTGTGGCCGAGCATGACCGAGGTGC 1863
Db      1582  GGTGGCCGTGTGAGACAGAGGTGCTTCCGTGTGTGGCCGAGCATGACCGAGGTGC 1641
Qy      1864  TTTGTGTGAGAGACAGAGGTGCTTGGGGGGGCCCTTGACCTTTGATGAAACAGATG 1923
Db      1642  TTTGTGTGAGAGACAGAGGTGCTTGGGGGGGCCCTTGACCTTTGATGAAACAGATG 1701
Qy      1924  GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1983
Db      1702  GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1761
Qy      1984  GAGCGAGAGATCGGCCCTACTAG 2007
Db      1762  GAGCGAGAGATCGGCCCTACTAG 1785

```

```

RESULT 5
US-09-949-016-1954
; Sequence 1954, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: C1001307

```

```

; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1954
; LENGTH: 1783
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-1954

Query Match      70.6%; Score 1429.6; DB 4; Length 1783;
Best Local Similarity 85.8%; Pred. No. 0;
Matches 1719; Conservative 0; Mismatches 64; Indels 221; Gaps 3;

Qy      4  GACCTTTGAGAGAACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 63
Db      1  GACCTTTGAGAGAACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 60
Qy      64  AGCCAAGCTGACAGCGAGAGAGGCTACAGTGTCTTACACCGCCAGCCCACTCAAGATAT 123
Db      61  AGCCAAGCTGACAGCGAGAGAGGCTACAGTGTCTTACACCGCCAGCCCACTCAAGATAT 120
Qy      124  GCAAGAGCAACCCAGGCAATATGAGCAACAAAGCTATGAACTTATGACAGCCCACTGAT 183
Db      121  GCAAGAGCAACCCAGGCAATATGAGCAACAAAGCTATGAACTTATGACAGCCCACTGAT 180
Qy      184  GTGAGTATATCCAGGCTGACAGCACTGCAACCTATAGGGAGAGCGGCTATGCACTTCT 243
Db      181  GTGAGTATATCCAGGCTGACAGCACTGCAACCTATAGGGAGAGCGGCTATGCACTTCT 240
Qy      244  TATGACAGCTTCCACTGTTTATCTACTCCACTGCCCCAGGCACTACAGAGCT 303
Db      241  TATGACAGCTTCCACTGTTTATCTACTCCACTGCCCCAGGCACTACAGAGCT 300
Qy      304  GTCCAGGGGTATGAGCACTGGTGTATGATACCACTGCTACAGTCAACCACTCAG 363
Db      301  GTCCAGGGGTATGAGCACTGGTGTATGATACCACTGCTACAGTCAACCACTCAG 360
Qy      364  GCCTCTTATGAGCTAGTCTGATATGAGCACTGAGCTGCTATCAGGCTATGGGAG 423
Db      361  GCCTCTTATGAGCTAGTCTGATATGAGCACTGAGCTGCTATCAGGCTATGGGAG 420
Qy      424  CAGCCAGAGCCACTGCACTACCAAGCCGAGATGAAACAAAGCCCACTGAGACTAGT 483
Db      421  CAGCCAGAGCCACTGCACTACCAAGCCGAGATGAAACAAAGCCCACTGAGACTAGT 480
Qy      484  CAATCTCATCTAGCAAGGGGTTACAAACAGCCCAAGCTTAGATATGAGAGATGAC 543
Db      481  CAATCTCATCTAGCAAGGGGTTACAAACAGCCCAAGCTTAGATATGAGAGATGAC 540
Qy      544  TAGAGTATCCAGGTAAGTCTGAGAGTACCCAGTACGAGCACTGCACTGCACTCCATCC 603
Db      541  TAGAGTATCCAGGTAAGTCTGAGAGTACCCAGTACGAGCACTGCACTGCACTCCATCC 600
Qy      604  TACCTCTTACAGACTATCTCTTACACAGCCGAGCTAGTATGATGAGAGACTTACT 663
Db      601  TACCTCTTACAGACTATCTCTTACACAGCCGAGCTAGTATGATGAGAGACTTACT 660
Qy      664  CAGCAGAAACCTATGAGGCAACGAGCAGCTATGACAGAGATGACTATGCTAACAA 723
Db      661  CAGCAGAAACCTATGAGGCAACGAGCAGCTATGACAGAGATGACTATGCTAACAA 720
Qy      724  AGAGCTATGAGGAGAGAGCTCCCACTAGTATACCAACCCCACTGATCTACAGCAA 783
Db      721  AGAGCTATGAGGAGAGAGCTCCCACTAGTATACCAACCCCACTGATCTACAGCAA 780
Qy      784  GTCCAAATATATAGCAACAGAGCAGCACTAGGGGAGAGAGTTCATTCAGAG 843

```





Dd		20	GACGCAAGTTAGAGAACAAGAGAAAGAGAGAAAATGGCGCTCACGAGATTACAGTACC	79
Qy		61	TATAGCCAAAGGTGACAGGCAGCAGGACTACAGTGCTTACACGCCAGGCCACTCAAGGA	120
Dd		80	TATAGCCAAGCTGACAGGACA-CAGGGCTTACAGTGTCTTACACGCCAGGCCACTCAAGGA	138
Qy		121	TATGACACAGAACCAACCAGGCATATGGGCAACAAAGCTATGGAACCTATGACAGGCCACT	180
Dd		139	TATGACACAGAACCAACCAGGCATATGGGCAACAAAGCTATGGAACCTATGACAGGCCACT	198
Qy		181	GATGTCAAGCTATACCCAGGCTCAGAACCATCTGCAACCTATGGGCAACCGCCCTATCAACT	240
Dd		199	GATGTCAAGCTATACCCAGGCTCAGAACCATCTGCAACCTATGGGCAACCGCCCTATGCAACT	258
Qy		241	TCTTATGAGACAGCCTCCCACTGGTTATTCATCTCCAACTGCCGCCAGGACTACAGCCAG	300
Dd		259	TCTTATGAGACAGCCTCCCACTGGTTATTCATCTCCAACTGCCGCCAGGACTACAGCCAG	318
Qy		301	CCTGTCCAGGGGATATGGCACTGTGTCTTATGATAACACACTGTCTACAGTCCACCACC	360
Dd		319	CCTGTCCAGGGGATATGGCACTGTGTCTTATGATAACACACTGTCTACAGTCCACCACC	378
Qy		361	CAGGCTCTTATGCAAGCTCACTTGCATATGGCACTCAGCCTGCTTATTCACACCTATGGG	420
Dd		379	CAGGCTCTTATGCAAGCTCACTTGCATATGGCACTCAGCCTGCTTATTCACACCTATGGG	438
Qy		421	CAGCAGCCAGCAGGCCA	436
Dd		439	CAGCAGCCAGCAGGCCA	454

RESULT 8  
 US-09-949-016-66382/c  
 ; Sequence 66382, Application US/09949016  
 ; Patent No. 6812339  
 ; GENERAL INFORMATION:  
 ; APPLICANT: VENTER, J. Craig et al.  
 ; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
 ; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
 ; FILE REFERENCE: C7001307

```

; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 66382
;
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-66382

Query Match      18.0%; Score 364; DB 4; Length 601;
Best Local Similarity 98.1%; Pred. No. 4,3e-88;
Matches 367; Conservative 1; Mismatches 6; Indels 0; Gaps 0

4  GGACGTTAGAGAAAGAGAGAGAAAGAGAAATATGCGCTCCACGGATTACAGTACCTAT 63
374  GGACCTTAGAGAAAGAGAGAGAAAGAGAAATATGCAATCCATGGAATTACAGTACCTAT 315
64  AGCCAGCTGCAGCCGACGAGGGCTTACAGTGTCTTACACCGCCAGGCCACTCAAGGATAT 123
314  AGCCAGCTGCAGCGAGGAGGGGCTTACAGTGTCTTACACCGCCAGGCCACTCAAGGATAT 255
124  GCACGAGCCACCCAGGCGCATATGGGCAACAAAGCTATGGAACCTATGACAGCCCACTGAT 183
254  GCACGAGCCACCCAGGCGCATATGGGCAACAAAGCTATGGAACCTATGACAGCCCACTGAT 195
154  GTCACTATATCCCAAGGCTCAGACCACTGCAACTATATGGGCAAGCCGCGCTATGCAACTTCT 243

```

1	2025	100.0	2390	18	US-10-791-017A-1	Sequence 1, Appl1
2	2023.4	99.9	2390	9	US-09-880-107-3769	Sequence 3769, Ap
3	2023.4	99.9	2390	10	US-09-960-705-1081	Sequence 1081, Ap
4	2023.4	99.9	2390	10	US-09-873-319-717	Sequence 717, Appl
5	1866.4	98.9	2376	9	US-09-822-830A-49	Sequence 49, Appl
6	1856.6	91.7	2273	9	US-09-882-830A-110	Sequence 410, App
7	1652.4	81.6	1988	17	US-10-094-749-1074	Sequence 1074, App
8	631	31.2	763	18	US-10-425-115-18296	Sequence 18296, Ap
9	570	28.1	2299	14	US-10-128-846-8847	Sequence 8847, Appl
10	514.4	25.4	540	9	US-09-864-761-8001	Sequence 8001, App
11	508.4	23.1	550	10	US-09-918-995-57650	Sequence 27690, Ap

C	12	481.6	23.7	568	10	US-09-98-86-795-5557	Sequence 9557, App
C	13	411.2	20.3	505	9	US-09-86-761-7766	Sequence 7766, App
C	14	399.4	19.7	401	10	US-09-91-88-995-1868	Sequence 3868, App
C	15	384.6	18.0	399	9	US-09-86-761-24705	Sequence 24705, App
C	16	370.6	19.3	417	9	US-09-96-351-10259	Sequence 10259, App
C	17	309	15.3	545	9	US-09-92-301-669	Sequence 669, App
C	18	269.8	13.3	361	9	US-09-96-335-3851	Sequence 3851, App
C	19	267.4	13.2	521	9	US-09-93-797-481	Sequence 481, App
C	20	258.4	12.8	503	9	US-09-86-761-8728	Sequence 8728, App
C	21	258.4	12.8	503	16	US-10-02-9-86-463	Sequence 4463, App
C	22	254.2	12.6	262	16	US-10-02-9-86-18163	Sequence 18163, App
C	23	253	12.5	253	9	US-09-86-761-25438	Sequence 25438, App
C	24	216.4	10.7	447	18	US-10-46-285-488	Sequence 488, App
C	25	188	9.3	459	9	US-09-86-761-2572	Sequence 2572, App
C	26	188	9.3	538	16	US-10-02-9-86-4273	Sequence 4273, App
C	27	187	9.2	188	16	US-10-02-9-86-17973	Sequence 17973, App
C	28	181	8.9	1256	18	US-10-72-3-860-8052	Sequence 8052, App
C	29	181	8.9	3132	13	US-10-02-7-632-111655	Sequence 111655, App
C	30	181	8.9	3132	13	US-10-02-7-632-114301	Sequence 114301, App
C	31	181	8.9	3132	17	US-10-02-7-632-111655	Sequence 111655, App
C	32	181	8.9	3132	17	US-10-02-7-632-114301	Sequence 114301, App
C	33	175.2	8.7	1822	18	US-10-75-5-889-649	Sequence 649, App
C	34	175.2	8.7	1824	17	US-10-43-9-703-58	Sequence 58, App
C	35	175.2	8.7	1939	10	US-09-91-9-39-122	Sequence 32, App
C	36	174.4	8.6	468	10	US-09-91-8-995-19709	Sequence 19709, App
C	37	144	7.1	187	9	US-09-86-761-19297	Sequence 19297, App
C	38	131.8	6.5	483	10	US-09-91-8-995-1865	Sequence 2865, App
C	39	129.6	6.4	147	9	US-09-86-761-24470	Sequence 24470, App
C	40	129.4	6.4	43712	18	US-10-72-3-860-1576	Sequence 1576, App
C	41	128	6.3	386	15	US-10-10-1-510-297	Sequence 297, App
C	42	116.4	5.7	450	9	US-09-738-973-145	Sequence 145, App
C	43	116.4	5.7	450	9	US-09-854-133-145	Sequence 145, App
C	44	116.4	5.7	450	15	US-10-14-649A-145	Sequence 145, App
C	45	100.4	5.0	460	10	US-09-91-8-995-17342	Sequence 27342, App

## ALIGNMENTS

```

RESULT 1
US-10-791-017A-1
? Sequence 1, Application US/10791017A
? Publication NO. US20060197827A1
? GENERAL INFORMATION:
? APPLICANT: JENAPHARM GmbH & Co. KG
? TITLE OF INVENTION: Methods for Determining Hormonal Effects of Substances
? FILE REFERENCE: Pat 3684/11
? CURRENT APPLICATION NUMBER: US/10/791,017A
? CURRENT FILING DATE: 2004-03-02
? NUMBER OF SEQ ID NOS: 7
? SOFTWARE: PatentIn Ver. 2.1
? SEQ ID NO 1
? LENGTH: 2390
? TYPE: DNA
? ORGANISM: Homo sapiens
? FEATURE:
? NAME/KEY: CDS
? LOCATION: (44)..(2011)
? OTHER INFORMATION: EMS
? US-10-791-017A-1

```

100.0%; Score 2025; DB 18; Length 2390;

ive 0; Mismatches 0; Indels 0; Gaps 0;

AAGAGAACGAGGAGGAGAGAGAAATGGCGTCCACGGATTACAGTACC 60

AGAGAACGAGGAGGAGAGAAATGCGTCCACGGATTACAGTACC 67

TTGCAGCGCAGCAGGGCTTACAGTGTACACCGCCAGCCCACTCAAGGA 120

TTGACGGCAGCAGGGCTACAGTGCTTACACCGCCAGCCCACTCAAGGA 127

QY 121 TATGACAGACCAACCGGATATGGGCAACAAAGCTATGGAACTTATGGACAGCCACT 180  
DB 128 TATGACAGACCAACCGGATATGGGCAACAAAGCTATGGAACTTATGGACAGCCACT 187  
QY 181 GATGTGAGCTATACCCAGGCTCAGACCACTGCAACCTATGGGAGACCGGCTATGGAACT 240  
DB 188 GATGTGAGCTATACCCAGGCTCAGACCACTGCAACCTATGGGAGACCGGCTATGGAACT 247  
QY 241 TCTTATGAGACGCTCCCACTGGTTATATCTCAACCTGCCCCAGGCAATACAGCTAG 300  
DB 248 TCTTATGAGACGCTCCCACTGGTTATATCTCAACCTGCCCCAGGCAATACAGCTAG 307  
QY 301 CCTGTCCAGGGGATATGGCACTGGTGTATGATACCACTGCTATACAGTACCAAGCACC 360  
DB 308 CCTGTCCAGGGGATATGGCACTGGTGTATGATACCACTGCTATACAGTACCAAGCACC 367  
QY 361 CAGGCTCTTATGAGCTCAGTGTGCAATATGGCACTGAGCTGTATTCAGGCTATGGG 420  
DB 368 CAGGCTCTTATGAGCTCAGTGTGCAATATGGCACTGAGCTGTATTCAGGCTATGGG 427  
QY 421 CAGGACCCAGACGCTCAGTGTGCAATATGGCACTGAGCTGTATTCAGGCTATGGG 480  
DB 428 CAGGACCCAGACGCTCAGTGTGCAATATGGCACTGAGCTGTATTCAGGCTATGGG 487  
QY 481 AGTCAACCTGATCTAGCAAGAGGGGTTTACAAACAGCCAGGCTTATGGATATGGACAGT 540  
DB 488 AGTCAACCTGATCTAGCAAGAGGGGTTTACAAACAGCCAGGCTTATGGATATGGACAGT 547  
QY 541 AACTACAGTTATCCCAAGTTACCTGGAGCTACCCCACTGCAAGGCTCAGTGTGCACTGCA 600  
DB 548 AACTACAGTTATCCCAAGTTACCTGGAGCTACCCCACTGCAAGGCTCAGTGTGCACTGCA 607  
QY 601 TCTTACCTCTTACAGCTATCTCTCTTACCAAGGCTATGTTATGATACAGAGTATAC 660  
DB 608 TCTTACCTCTTACAGCTATCTCTCTTACCAAGGCTATGTTATGATACAGAGTATAC 667  
QY 661 TCTCAGAGAAACCTATGGGCAACCGAGGCTATGAGCAAGGCTATGAGTATGAGTAC 720  
DB 668 TCTCAGAGAAACCTATGGGCAACCGAGGCTATGAGCAAGGCTATGAGTATGAGTAC 727  
QY 721 CAAAGCAGTATGGGCAAGGCTCTCCCACTATGTTACCCCACTGATGATCTTACAGC 780  
DB 728 CAAAGCAGTATGGGCAAGGCTCTCCCACTATGTTACCCCACTGATGATCTTACAGC 787  
QY 781 CAAAGCTCAAGTCAATATGAGCAACAGAGCAGGCTACCGGCAAGGCTATGATCTTAC 840  
DB 788 CAAAGCTCAAGTCAATATGAGCAACAGAGCAGGCTACCGGCAAGGCTATGATCTTAC 847  
QY 841 CAGGACCAACCCAGTATGAGTGTATGAGGAGGCTTATGAGGATTTTCCGAGCA 900  
DB 848 CAGGACCAACCCAGTATGAGTGTATGAGGAGGCTTATGAGGATTTTCCGAGCA 907  
QY 901 GAGAGAAACCGGACATGAGTGGCTGTATTAACCGGGGCAAGGGGAGAGGAGGATTTGAT 960  
DB 908 GAGAGAAACCGGACATGAGTGGCTGTATTAACCGGGGCAAGGGGAGAGGAGGATTTGAT 967  
QY 961 CTTGAGAGGCTATGAGGAGGCTGGGAGGAGAGGAGGAGGAGGAGGAGGAGGAGGCTGGA 1020  
DB 968 CTTGAGAGGCTATGAGGAGGCTGGGAGGAGAGGAGGAGGAGGAGGAGGAGGAGGCTGGA 1027  
QY 1021 GAGGAGGAGGCTTCAATTAAGCTGTGAGACCACTGATGAGGAGGAGGAGGAGGAGGCTG 1080  
DB 1028 GAGGAGGAGGCTTCAATTAAGCTGTGAGACCACTGATGAGGAGGAGGAGGAGGAGGCTG 1087  
QY 1081 GAGGACCTGTATGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGATTA 1140  
DB 1088 GAGGACCTGTATGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGATTA 1147  
QY 1141 AATGACAGTGTGATCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1200  
DB 1148 AATGACAGTGTGATCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1207  
QY 1201 ATGAACAAGAGAACTGGGCAACCATGATGATGATGATGATGATGATGATGATGATGATG 1260

DB 1208 ATGAACAAGAGAACTGGGCAACCATGATGATGATGATGATGATGATGATGATGATGATG 1267  
QY 1261 CCCAAAGGAGATGACCAAGTGTCTTATGAAGACCCCACTGCAAGGCTGCGTGGAA 1320  
DB 1268 CCCAAAGGAGATGACCAAGTGTCTTATGAAGACCCCACTGCAAGGCTGCGTGGAA 1327  
QY 1321 TGGTTGATGAGGAGAAATTTTCAAGGAGCAAACTTAAAGTCTCCCTTGTCCGAGAAAG 1380  
DB 1328 TGGTTGATGAGGAGAAATTTTCAAGGAGCAAACTTAAAGTCTCCCTTGTCCGAGAAAG 1387  
QY 1381 CCTTCATGAACAGTATGAGGAGGCTGTCTGCAACCCCGTGAAGGAGGAGGATGCCA 1440  
DB 1388 CTTCAATGAACAGTATGAGGAGGCTGTCTGCAACCCCGTGAAGGAGGAGGATGCCA 1447  
QY 1441 CCACTCCGTGAGAGTCCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1500  
DB 1448 CCACTCCGTGAGAGTCCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1507  
QY 1501 GGCCTGAGAGGAGATGAGGAGGCTTCCCTCCAAAGAGGAGGAGGAGGAGGAGGAGGAG 1560  
DB 1508 GGCCTGAGAGGAGATGAGGAGGCTTCCCTCCAAAGAGGAGGAGGAGGAGGAGGAGGAG 1567  
QY 1561 CCCTTGAAGAGGAGAAACGTCAGCAACGAGCTGAGAGTGTGCAAGTGTCCCAATCCGGGT 1620  
DB 1568 CCCTTGAAGAGGAGAAACGTCAGCAACGAGCTGAGAGTGTGCAAGTGTCCCAATCCGGGT 1627  
QY 1621 TGTGAAACCAAACTTCCGCTGAGAAACAGAGTGTGCAACCAAGTGTAAAGGCTCCAAAGGCT 1680  
DB 1628 TGTGAAACCAAACTTCCGCTGAGAAACAGAGTGTGCAACCAAGTGTAAAGGCTCCAAAGGCT 1687  
QY 1681 GAGAGCTTCTCCGACCACTTCCGACCCCGGAGTGTATGATGATGATGATGATGATGATG 1740  
DB 1688 GAGAGCTTCTCCGACCACTTCCGACCCCGGAGTGTATGATGATGATGATGATGATGATG 1747  
QY 1741 GGTGCAATGAGGAGGAGAAAGGCTGCTCATGATGATGATGATGATGATGATGATGATG 1800  
DB 1748 GGTGCAATGAGGAGGAGAAAGGCTGCTCATGATGATGATGATGATGATGATGATGATG 1807  
QY 1801 AGAGTGTGCTGTGTGAGAGCAAGGCTTCCGCTGTGTGTGTGTGTGTGTGTGTGTGTGT 1860  
DB 1808 AGAGTGTGCTGTGTGAGAGCAAGGCTTCCGCTGTGTGTGTGTGTGTGTGTGTGTGTGT 1867  
QY 1861 GGCCTTGTGTGAGAGAAACAGAGTGTGCTTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1920  
DB 1868 GGCCTTGTGTGAGAGAAACAGAGTGTGCTTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1927  
QY 1921 ATGGAGAGAAAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1980  
DB 1928 ATGGAGAGAAAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1987  
QY 1981 CAGAGGCGCAGAGATGAGGCTTCTATGATGAGAGAGGAGGAGGAGGAGGAGGAGGAG 2025  
DB 1988 CAGAGGCGCAGAGATGAGGCTTCTATGATGAGAGAGGAGGAGGAGGAGGAGGAGGAG 2032

RESULT 2  
US-09-880-107-3769  
Sequence 3769, Application US/09880107  
Patent No. US20020142981A1  
GENERAL INFORMATION:  
APPLICANT: Horne, Darci T.  
APPLICANT: Vockley, Joseph G.  
APPLICANT: Scheif, Uwe  
APPLICANT: Gene Logic, Inc.  
TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer  
FILE REFERENCE: 44921-5028-WO  
CURRENT APPLICATION NUMBER: US/09/880,107  
CURRENT FILING DATE: 2001-06-14  
PRIOR APPLICATION NUMBER: US 60/211,379  
PRIOR FILING DATE: 2000-06-14  
PRIOR APPLICATION NUMBER: US 60/237,054  
PRIOR FILING DATE: 2000-10-02







Qy	1681	GAAGGCTTCCTCCGCGCACCCCTTTCCGCGCCCGGGGTGGATCGTGGAGAGGTGGCCCT	1740
Db	1688	GAAGGCTTCCTCCGCGCACCCCTTTCCGCGCCCGGGGTGGATCGTGGAGAGGTGGCCCT	1747
Qy	1741	GGTGGCATGCGGGAGAGAAAGGTGGCTCATGTGATCGTGTGTCTCCGGTGAATGTTTC	1800
Db	1748	GOTGGCATGCGGGAGAGAAAGGTGGCTCATGTGATCGTGTGTGTCTCCGGTGAATGTTTC	1807
Qy	1801	AGAGGTGCGCGTGTGGAGACAGAGGTGGCTTCCGTGTGTGCGCGGGCATGAGCCGAGGT	1866
Db	1808	AGAGGTGCGCGTGTGGAGACAGAGGTGGCTTCCGTGTGTGCGCGGGCATGAGCCGAGGT	1867
Qy	1861	GGCTTTGGTGGAGAGAAAGAGGTGGCCCTGTGGGGGGCCCCCTGGACCTTTGATGGAAAG	1922
Db	1868	GGCTTTGGTGGAGAGAAAGAGGTGGCCCTGTGGGGGGCCCCCTGGACCTTTGATGGAAAG	1927
Qy	1921	ATGGAGAGAAAG	1980
Db	1928	ATGGAGAGAAAG	1987
Qy	1981	CAGGAGCGCAGAGATCGGCCCTTACTGATGACAGAGACCCCGCAGA	2025
Db	1988	CAGGAGCGCAGAGATCGGCCCTTACTGATGACAGAGACCCCGCAGA	2032
RESULT 4			
US-09-873-319-717			
Sequence 717, Application US/09873319A			
Publication No. US20030134324A1			
GENERAL INFORMATION:			
APPLICANT: Mungert, William E.			
APPLICANT: Kulkarni, Prakash			
APPLICANT: Getzenberg, Robert H.			
APPLICANT: Waga, Iwao			
APPLICANT: Yamamoto, Jun			
TITLE OF INVENTION: Identifying Drugs for and Diagnosis of Benign Prostatic			
TITLE OF INVENTION: Hyperplasia Using Gene Expression Profiles			
FILE REFERENCES: 44921-5029-US			
CURRENT FILING DATE: 2001-06-05			
EARLIER APPLICATION NUMBER: US 60/223,323			
EARLIER FILING DATE: 2000-08-07			
NUMBER OF SEQ ID NOS: 755			
SOFTWARE: PatentIn Ver. 2.1			
SEQ ID NO 717			
LENGTH: 2390			
TYPE: DNA			
ORGANISM: Homo sapiens			
FEATURE:			
OTHER INFORMATION: Genbank Accession No. US20030134324A1 X66899			
US-09-873-319-717			
Query Match			
Best Local Similarity 99.9%; Score 2023.4; DB 10; Length 2390;			
Matches 2024; Conservative 0; Mismatches 1; Indels 0; Gaps 0;			
Qy	1	GACCGACCTTGAG	60
Db	8	GACCGACCTTGAG	67
Qy	61	TATAGCCAGCTGCAGCGCAGAGGGCTACAGTGTCTTACACCGCCAGCCCACTTCAAGA	120
Db	68	TATAGCCAGCTGCAGCGCAGAGGGCTACAGTGTCTTACACCGCCAGCCCACTTCAAGA	127
Qy	121	TATGACAGACCAACCGAGCATATGGGGCAACAAAGCTATGGAACTTATGAGACGCCACT	180
Db	128	TATGACAGACCAACCGAGCATATGGGGCAACAAAGCTATGGAACTTATGAGACGCCACT	187
Qy	181	GATGTCAGCTATACCGAGGCTCAGACCACTGCAACCTATGGGAGACCGCTATGCAACT	240
Db	188	GATGTCAGCTATACCGAGGCTCAGACCACTGCAACCTATGGGAGACCGCTATGCAACT	247
Qy	241	TCTTATGAGACCTCCCACTGGTTATATCTACTCCACTGCCCGCCAGGCAATACAGCCAG	300

Db	248	TCCTTATGGACAGCTCCCACTGGTTATATCTCAACTGCCCCAGGCACTAAGCCAG	307
Oy	301	CTGTCCAGGGGTATGGCACTGGTGTCTTATGATACACACTGCTACAGTCAACCAACC	360
Db	308	CCTGTCCAGGGGTATGGCACTGGTGTCTTATGATACACACTGCTACAGTCAACCAACC	367
Oy	361	CAGGCTTCTATGCAAGCTCAAGTCTGCAATATGGCACTAGAGCTGCTTATCCAGCTATGGG	420
Db	368	CAGGCTTCTATGCAAGCTCAAGTCTGCAATATGGCACTAGAGCTGCTTATCCAGCTATGGG	427
Oy	421	CAGAGCCAGAGGCCACTGCACTTACAAAGCCGAGAGTGGAAACAAGCCCACTGAGACT	480
Db	428	CAGAGCCAGAGGCCACTGCACTTACAAAGCCGAGAGTGGAAACAAGCCCACTGAGACT	487
Oy	481	AGTCAACTCTCAATCTAGACAAGGGGGTTACAAACAGCCCAAGCTTAGATATGACAAGT	540
Db	488	AGTCAACTCTCAATCTAGACAAGGGGGTTACAAACAGCCCAAGCTTAGATATGACAAGT	547
Oy	541	AACCTACAGTTATCCCAAGTACTCTGGAGACTACCCCATGCAAGCCAGTCACTGCCA	600
Db	548	AACCTACAGTTATCCCAAGTACTCTGGAGACTACCCCATGCAAGCCAGTCACTGCCA	607
Oy	601	TCCTAACCTCCTACAGCTATTCCTCTACACAGCCGACTAGTTATGATCAGAGAGTTAC	660
Db	608	TCCTAACCTCCTACAGCTATTCCTCTACACAGCCGACTAGTTATGATCAGAGAGTTAC	667
Oy	661	TCTCAGAGAAACACTATATGGGCAACCGAGCAGTATGAGACAGAGTAGCTATGTCAA	720
Db	668	TCTCAGAGAAACACTATATGGGCAACCGAGCAGTATGAGACAGAGTAGCTATGTCAA	727
Oy	721	CAAGCAGACTATGGGCAAGAGCTCCCACTAGTTACCAACCCAACTGATCTCAAGC	780
Db	728	CAAGCAGACTATGGGCAAGAGCTCCCACTAGTTACCAACCCAACTGATCTCAAGC	787
Oy	781	CAAGCTCAAGTCAATATAGCCACAAGAGCAGAGCTACGGGCGCAGAGTTCAATTCCGA	840
Db	788	CAAGCTCAAGTCAATATAGCCACAAGAGCAGAGCTACGGGCGCAGAGTTCAATTCCGA	847
Oy	841	CAGGACCAACCCCACTAGCACTGGGTGTTTATATGGGCAAGTCTGAGAGATTTCGCGACA	900
Db	848	CAGGACCAACCCCACTAGCACTGGGTGTTTATATGGGCAAGTCTGAGAGATTTCGCGACA	907
Oy	901	GGAGAGAACCGGACATGAGTGCGCTTGATTAACCGGGCGAGGGGAAGAGGGGATTTGAT	960
Db	908	GGAGAGAACCGGACATGAGTGCGCTTGATTAACCGGGCGAGGGGAAGAGGGGATTTGAT	967
Oy	961	CGTGGAGGCATGAGCAGAGGTGGCGGGGAGAGAGACCGGTGGAATGGGCAAGCGCTGGA	1020
Db	968	CGTGGAGGCATGAGCAGAGGTGGCGGGGAGAGAGACCGGTGGAATGGGCAAGCGCTGGA	1027
Oy	1021	GAGCGAGGTGGCTTCAATTAAGCCGTGGTGGACCCATGGAATGAAGGACCAAGATCTTGATCTA	1080
Db	1028	GAGCGAGGTGGCTTCAATTAAGCCGTGGTGGACCCATGGAATGAAGGACCAAGATCTTGATCTA	1087
Oy	1081	GGCCCACTGTAGATCCAGATGGAAGACTCTGACAACAGTGCATTTATGTACAAAGATTAA	1140
Db	1088	GGCCCTCTGTAGATCCAGATGGAAGACTCTGACAACAGTGCATTTATGTACAAAGATTAA	1147
Oy	1141	AATGACAGTGTGACTCTAGATGATCTGGCAGACTTCTTTAAGCAGTGTGGGTGTTAAG	1200
Db	1148	AATGACAGTGTGACTCTAGATGATCTGGCAGACTTCTTTAAGCAGTGTGGGTGTTAAG	1207
Oy	1201	ATGAAACAAGAAATGGGGCAACCATATATCAATCACTGACCAAGGAAACAAGAAAG	1260
Db	1208	ATGAAACAAGAAATGGGGCAACCATATATCAATCACTGACCAAGGAAACAAGAAAG	1267
Oy	1261	CCCAAAAGGCATGCAAGTGTCTTATGAAGACCCCACTGCCAAGGCTGCGGTGAA	1320
Db	1268	CCCAAAAGGCATGCAAGTGTCTTATGAAGACCCCACTGCCAAGGCTGCGGTGAA	1327
Oy	1321	TGGTTGATGGGAAGAATTTTCAAGGGAGCAAACTTAAGTCTCCCTTGTGCGAAGAG	1380

Db	1128	TGCTTGAATGGGAAAGATTTTCAAGGAGCAAACTTAAAGTCTCCCTTGTCTGGAAAG	1387
Oy	1381	CCTCCATGAAACATGATACGGGGTGTGTTCACACCCGTGAGGGCAGAGGCATGTCCACCA	1440
Db	1388	CCTCCAAAGAACAGATATCGGGGTGTGTTCACACCCGTGAGGGCAGAGGCATGTCCACCA	1447
Oy	1441	CCACTCCGTGAGGCTCAAGAGGCCCAAGAGGTCTCTGGAGAGCCATGGTGTGCAATGGGA	1500
Db	1448	CCACTCCGTGAGGCTCAAGAGGCCCAAGAGGTCTCTGGAGAGCCATGGTGTGCAATGGGA	1507
Oy	1501	GGCGGTGAGGAGATAGAGAGGCTTCCCTCCAAAGAGACCCCGGGGTTTCCGAGGGAAAC	1560
Db	1508	GGCGGTGAGGAGATAGAGAGGCTTCCCTCCAAAGAGACCCCGGGGTTTCCGAGGGAAAC	1567
Oy	1561	CCCTCTGGAGAGGAAAGTCCAGAGCCAGAGCTGGAGCTGGAGTGTCCCAATCCGGGT	1620
Db	1568	CCCTCTGGAGAGGAAAGTCCAGAGCCAGAGCTGGAGCTGGAGTGTCCCAATCCGGGT	1627
Oy	1621	TGTGAAACCAAGAACTTGCCCTTGAGAAACAAGTGTCAACAAGTGTAAAGGCCCAAGCTT	1680
Db	1628	TGTGAAACCAAGAACTTGCCCTTGAGAAACAAGTGTCAACAAGTGTAAAGGCCCAAGCTT	1687
Oy	1681	GAAGGCTTCTTCCCGCCACCTTTTCCGGCCCCCGGTGTGATTCGTGCAAGAGTGGCCCT	1740
Db	1688	GAAGGCTTCTTCCCGCCACCTTTTCCGGCCCCCGGTGTGATTCGTGCAAGAGTGGCCCT	1747
Oy	1741	GATGGCATGCGGGGAGGAAGAGTGGCTCTCATGATTCGTGTGTGTCCCGTGTGAATGTTC	1800
Db	1748	GATGGCATGCGGGGAGGAAGAGTGGCTCTCATGATTCGTGTGTGTCCCGTGTGAATGTTC	1807
Oy	1801	AGAGGTGGCGCTGTGTGTGAGACAGAGGTGGCTTCGTGTGTGTCCCGGGCAATGGAACGAGT	1860
Db	1808	AGAGGTGGCGCTGTGTGTGAGACAGAGGTGGCTTCGTGTGTGTCCCGGGCAATGGAACGAGT	1867
Oy	1861	GGCTTTGTGTGAGAGAAACAAGAGTGGCCCTTGAGGGGCCCCCTTGACCTTTGATGGAAACAG	1920
Db	1868	GGCTTTGTGTGAGAGAAACAAGAGTGGCCCTTGAGGGGCCCCCTTGACCTTTGATGGAAACAG	1927
Oy	1921	ATGGAGAGAAAGAAAGAGAGAGCTGTGAGAGACTGTGAAAAAATGATTAAGCGAGCACCGT	1980
Db	1928	ATGGAGAGAAAGAAAGAGAGAGCTGTGAGAGACTGTGAAAAAATGATTAAGCGAGCACCGT	1987
Oy	1981	CAGAGCGCAGAGATCCGCGCTCACTAAGATGCAAGAAACCCCGGAGA	2025
Db	1988	CAGAGCGCAGAGATCCGCGCTCACTAAGATGCAAGAAACCCCGGAGA	2032

```

/ RESULT 5
/ US-09-822-830A-49/c
/ Sequence 49, Application US/09822830A
/ Patent No. US20020142952A1
/ GENERAL INFORMATION:
/ APPLICANT: Genetics Institute, Inc.
/ APPLICANT: Wong, Gordon G.
/ APPLICANT: Clark, Hilary
/ APPLICANT: Fechtel, Kim
/ APPLICANT: Agostino, Michael J.
/ APPLICANT: Howes, Steven H.
/ APPLICANT: Resnick, Richard J.
/ APPLICANT: Gulukota, Kamalakara
/ APPLICANT: Graham, James R.
/ TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING NOVEL, SECRETED PROTEINS
/ FILE REFERENCE: GIN 6402
/ CURRENT APPLICATION NUMBER: US/09/822,830A
/ CURRENT FILING DATE: 2001-03-29
/ PRIOR APPLICATION NUMBER: 60/195,604
/ PRIOR FILING DATE: 2000-04-06
/ NUMBER OF SEQ ID NOS: 631
/ SOFTWARE: PatentIn Ver. 2.0
/ SEQ ID NO 49
/ LENGTH: 2176
/ TYPE: DNA
/ ORGANISM: Homo sapiens

```

[illegible]

1046 GTGAGCCATGATGAGAGCAAGATCTGATCTAGGAGCCACCTGTAGATCCAGATGAG 1105  
1152 GTGAGCCATGATGAGAGCAAGATCTGATCTAGGAGCCACCTGTAGATCCAGATGAG 1093  
1106 ACTCTGACCAAGTGAATTTATGATGAGATTAATGACAGTGTGATCTGATGATG 1165  
1092 ACTCTGACCAAGTGAATTTATGATGAGATTAATGACAGTGTGATCTGATGATG 1033  
1166 TGGCAGATCTTTTAAAGCAGTGTGGGTGTTAAGATGAAAGAGAACTGGGCAACCA 1225  
1032 TGGCAGATCTTTTAAAGCAGTGTGGGTGTTAAGATGAAAGAGAACTGGGCAACCA 973  
1226 TGAATCCATCTCTGAG 1285  
972 TGAATCCATCTCTGAG 913  
1286 ATGAG 1345  
912 ATGAG 853  
1346 GAG 1405  
852 GAG 793  
1406 GTCTGCAACCCCTGAG 1465  
792 GTCTGCAACCCCTGAG 733  
1466 CAG 1525  
732 CAG 673  
1526 TCCCTCCAG 1585  
672 TCCCTCCAG 613  
1586 ACCGAGCTGAG 1645  
612 ACCGAGCTGAG 553  
1646 GAG 1705  
552 GAG 493  
1706 GCGCCCGAG 1765  
492 GCGCCCGAG 433  
1766 GCGCTCATGATCTGT 1825  
432 GCGCTCATGATCTGT 373  
1826 GTGAGCTTCTGT 1885  
372 GTGAGCTTCTGT 313  
1886 GCGCTGAG 1945  
312 GCGCTGAG 253  
1946 GAG 2005  
252 GAG 193  
2006 AGATGAG 2025  
192 AGATGAG 173

RESULT 6  
US-09-822-830A-410/c  
; Sequence 410, Application US/09822830A

Patent No. US20020142952A1  
; GENERAL INFORMATION:  
; APPLICANT: Genetics Institute, Inc.  
; APPLICANT: Wong, Gordon G.  
; APPLICANT: Clark, Hilary  
; APPLICANT: Fehchel, Kim  
; APPLICANT: Agostino, Michael J.  
; APPLICANT: Howes, Steven H.  
; APPLICANT: Resnick, Richard J.  
; APPLICANT: Gulukota, Kamalakari  
; APPLICANT: Graham, James R.  
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING NOVEL SECRETED PROTEINS  
; FILE REFERENCE: GIN 6402  
; CURRENT APPLICATION NUMBER: US/09/822, 830A  
; CURRENT FILING DATE: 2001-03-29  
; PRIOR APPLICATION NUMBER: 60/195, 604  
; PRIOR FILING DATE: 2000-04-06  
; NUMBER OF SEQ ID NOS: 631  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 410  
; LENGTH: 2273  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-822-830A-410

Query Match 91.7%; Score 1856.6; DB 9; Length 2273;  
Best Local Similarity 97.1%; Pred. No. 0;  
Matches 1933; Conservative 0; Mismatches 4; Indels 54; Gaps 2;  
Query 35 AATGAGCTGACAGATTAAGTACCTATAGCAAGCTGAGAGAGAGAGAGAGAGAG 94  
Db 2268 AAGTGTGCTGACAGATTAAGTACCTATAGCAAGCTGAGAGAGAGAGAGAGAG 2209  
Query 95 CTGACACCGCCAGCCCACTCAAGATATGACAGACCAAGAGATATGAGCAACAA 154  
Db 2208 CTGACACCGCCAGCCCACTCAAGATATGACAGACCAAGAGATATGAGCAACAA 2149  
Query 155 GCTATGAACTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 214  
Db 2148 GCTATGAACTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2089  
Query 215 CCTATGAG 274  
Db 2088 CCTATGAG 2029  
Query 275 CAATGCCCCCAGAGATACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 334  
Db 2028 CAATGCCCCCAGAGATACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1969  
Query 335 CCACTGCTGATGATCAACCAACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 394  
Db 1968 CCACTGCTGATGATCAACCAACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1909  
Query 395 CTCAGCTGCTTATCAAGCTATAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 454  
Db 1908 CTCAGCTGCTTATCAAGCTATAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1849  
Query 455 AGGATGAAACAG 514  
Db 1848 AGGATGAAACAG 1789  
Query 515 AGCCAGCTGATGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 574  
Db 1788 AGCCAGCTGATGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1729  
Query 575 CCAATGAG 634  
Db 1728 CCAATGAG 1669  
Query 635 CGACTAGTATGATCAG 694  
Db 1668 CGACTAGTATGATCAG 1609

QY 695 ATGACAGCAGAGTATGTCATCAAGCAGTATGAGGAGCAGCCTCCAGTATGTT 754  
DB 1608 ATGACAGCAGAGTATGTCATCAAGCAGTATGAGGAGCAGCCTCCAGTATGTT 1549  
QY 755 ACCACCCCAAGCTGATCTTACAGCCAGCCTCCAGTATGAGCAGCAGCAGCA 814  
DB 1548 ACCACCCCAAGCTGATCTTACAGCCAGCCTCCAGTATGAGCAGCAGCAGCA 1489  
QY 815 GCTACGGGAGCAGAGTTCATCCGACAGCAGCAGCAGCAGCAGCAGCAGCAGC 874  
DB 1488 GCTACGGGAGCAGAGTTCATCCGACAGCAGCAGCAGCAGCAGCAGCAGCAGC 1429  
QY 875 AGGAGTCTGAGAGTTCCTCCGACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 934  
DB 1428 AGGAGTCTGAGAGTTCCTCCGACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 1369  
QY 935 GGGGAGGAGAGAGGAGGAGTTCGATGAGGAGCAGCAGCAGCAGCAGCAGCAGC 994  
DB 1368 GGGGAGGAGAGAGGAGGAGTTCGATGAGGAGCAGCAGCAGCAGCAGCAGCAGC 1309  
QY 995 GAGCGGCTGAGATGAGGAGGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 1054  
DB 1308 GAGCGGCTGAGATGAGGAGGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 1252  
QY 1055 TGAGTGAAGAGCAGATCTTATGATGAGGAGCAGCAGCAGCAGCAGCAGCAGCAGC 1114  
DB 1251 TGAGTGAAGAGCAGATCTTATGATGAGGAGCAGCAGCAGCAGCAGCAGCAGCAGC 1192  
QY 1115 ACAAGTCAATTTATGATCAAGATTAATGACAGTGTGATGATGATGATGATGATGAT 1174  
DB 1191 ACAAGTCAATTTATGATCAAGATTAATGACAGTGTGATGATGATGATGATGATGAT 1132  
QY 1175 TCTTTAAGCAGTGTGGGCTGTTAAGTGAAGCAAGAGAGAGAGAGAGAGAGAGAGAG 1234  
DB 1131 TCTTTAAGCAGTGTGGGCTGTTAAGTGAAGCAAGAGAGAGAGAGAGAGAGAGAGAG 1072  
QY 1235 TCTAAGTGAAG 1294  
DB 1071 TCTAAGTGAAG 1012  
QY 1295 CACCACTGCGAAGGCTGCGGTGAGATGTTGATGAGAGAGAGAGAGAGAGAGAGAG 1354  
DB 1011 CACCACTGCGAAGGCTGCGGTGAGATGTTGATGAGAGAGAGAGAGAGAGAGAGAG 952  
QY 1355 TTTAAAGTCTCTCTGCTGCGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1414  
DB 951 TTTAAAGTCTCTCTGCTGCGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 892  
QY 1415 CCGGTGAG 1474  
DB 891 CCGGTGAG 832  
QY 1475 CTGGGAG 1534  
DB 831 CTGGGAG 772  
QY 1535 GAG 1594  
DB 771 GAG 712  
QY 1595 GAG 1654  
DB 711 GAG 652  
QY 1655 GCAACAGAGTGAAG 1714  
DB 651 GCAACAGAGTGAAG 643  
QY 1715 GTGGTGAATCTGAG 1774  
DB 642 GTGGTGAATCTGAG 583  
QY 1775 ATCGAGTGTGCTCGGCTGAG 1834

DB 582 ATCGAGTGTGCTCGGCTGAG 523  
QY 1835 GTGGTGAATCTGAG 1894  
DB 522 GTGGTGAATCTGAG 463  
QY 1895 GGGGCTGAG 1954  
DB 462 GGGGCTGAG 403  
QY 1955 GAAAAATGATTAAG 2014  
DB 402 GAAAAATGATTAAG 343  
QY 2015 GACCCCGCAGAG 2025  
DB 342 GACCCCGCAGAG 332

RESULT 7  
US-10-094-749-1074  
Sequence 1074, Application US/10094749  
Publication No. US20030219741A1  
GENERAL INFORMATION:  
APPLICANT: ISOGAI, TAKAO  
APPLICANT: SUGIYAMA, TOMOYASU  
APPLICANT: OTSUKI, TETSUJI  
APPLICANT: WAKAMATSU, AI  
APPLICANT: SATO, HIROYUKI  
APPLICANT: ISHII, SHIZUKO  
APPLICANT: YAMAMOTO, JUN-ICHI  
APPLICANT: ISONO, YUUKO  
APPLICANT: HIO, YUKI  
APPLICANT: OTSUKA, KAORU  
APPLICANT: NAGAI, KEIICHI  
APPLICANT: IRIE, RYOTARO  
APPLICANT: TAMECHIKI, ICHIRO  
APPLICANT: SEKI, NAOHITO  
APPLICANT: YOSHIKAWA, TSUTOMU  
APPLICANT: OTSUKA, MOTOTYUKI  
APPLICANT: NAGAHARI, KENJI  
APPLICANT: MASUHO, YASUHIKO  
TITLE OF INVENTION: NOVEL FULL-LENGTH cDNA  
FILE REFERENCE: 08435/0160  
CURRENT FILING DATE: 2002-03-12  
PRIORITY FILING DATE: 2002-01-24  
PRIORITY FILING DATE: 2001-09-14  
PRIORITY FILING DATE: 2001-09-14  
NUMBER OF SEQ ID NOS: 3381  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 1074  
LENGTH: 1988  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-10-094-749-1074

Query Match 81.6%; Score 1652.4; DB 17; Length 1988;  
Best Local Similarity 91.5%; Pred. No. 0;  
Matches 1831; Conservative 0; Mismatches 1; Indels 168; Gaps 1;

QY 26 AAG 85  
DB 1 AAG 60  
QY 86 GCTACAGTGTCTTACACCGCCAG 145  
DB 61 GCTACAGTGTCTTACACCGCCAG 120  
QY 146 GCGAACAAGCTATGAG 205

Db 121 GGCAACAAAGCTATGAAACCTATGAGAGCCCACTGATGCTATATCCAGGCTCAGA 180  
Qy 206 CCACTGCAACCTATGAGGAGAGCCCTTATGCACTTCTTATGACAGCTTCCACTGTT 265  
Db 181 CCACTGCAACCTATGAGGAGAGCCCTTATGCACTTCTTATGACAGCTTCCACTGTT 240  
Qy 266 ATACTACTGCACTGCCCCCAGGAGATACAGCCAGCTGTCAGGGGTAATGGGCACTGGT 325  
Db 241 ATACTACTGCACTGCCCCCAGGAGATACAGCCAGCTGTCAGGGGTAATGGGCACTGGT 300  
Qy 326 CTATATGATACCACTGCTACAGTCAACAACCAAGGCTCTTATGCACTGCTCAGTCTG 385  
Db 301 CTATATGATACCACTGCTACAGTCAACAACCAAGGCTCTTATGCACTGCTCAGTCTG 360  
Qy 386 CATATGCACTGAGCTGCTTATATCAGCTTATGAGGAGAGCCAGCCAGCCACTGCACTA 445  
Db 361 CATATGCACTGAGCTGCTTATATCAGCTTATGAGGAGAGCCAGCCAGCCACTGCACTA 420  
Qy 446 CAAGACCGCAAGATGAAAACAAGCCCACTGAGACTGAACTCAACTCAAGCAAGAGG 505  
Db 421 CA----- 422  
Qy 506 GTTACAACAGCCAGCCAGCTAGATATGAGACAGAGTAACTACAGTTATCCAGGTAACCTG 565  
Db 423 ----- 422  
Qy 566 GGAAGCTACCCCACTGAGCCAGTCACTGCACTCTCATCTTACCTTCAACAGCTATTTCT 625  
Db 423 -----AGCTATTTCT 432  
Qy 626 CTACACAGCCGACTAGTTATGATGAGAGAGTTACTCTGAGAGAAACCTATGAGGAGAC 685  
Db 433 CTACACAGCCGACTAGTTATGATGAGAGAGTTACTCTGAGAGAAACCTATGAGGAGAC 492  
Qy 686 CGAGCACTATGAGCAGCAGAGTATGATGATGATGATGATGATGATGATGATGATGATG 745  
Db 493 CGAGCACTATGAGCAGCAGAGTATGATGATGATGATGATGATGATGATGATGATGATG 552  
Qy 746 CCACTAGTTACCCCACTGAGATCTGATCTTCAAGCCAGCTCCAGTCAATATATGCAAC 805  
Db 553 CCACTAGTTACCCCACTGAGATCTTCAAGCCAGCTCCAGTCAATATATGCAAC 612  
Qy 806 AGAGCAGCAGCTTACGAGCAGAGCTTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 865  
Db 613 AGAGCAGCAGCTTACGAGCAGAGCTTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 672  
Qy 866 TTTATGAGCAGAGTCTGAGAGATTTTCCGAGCAGAGAGAGAGAGAGAGAGAGAGAG 925  
Db 673 TTTATGAGCAGAGTCTGAGAGATTTTCCGAGCAGAGAGAGAGAGAGAGAGAGAGAG 732  
Qy 926 CTGATTAACCGGAG 985  
Db 733 CTGATTAACCGGAG 792  
Qy 986 GGGAG 1045  
Db 793 GGGAG 852  
Qy 1046 GTGAGCCATGATGATGAG 1105  
Db 853 GTGAGCCATGATGATGAG 912  
Qy 1106 ACTCTGACAAAGTGAATTTATGATGATGATGATGATGATGATGATGATGATGATGAT 1165  
Db 913 ACTCTGACAAAGTGAATTTATGATGATGATGATGATGATGATGATGATGATGATGAT 972  
Qy 1166 TGGCAGACTCTTAAAGCAGTGGGTTTAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1225  
Db 973 TGGCAGACTCTTAAAGCAGTGGGTTTAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1032  
Qy 1226 TGATCCACATCTACTGAG 1285  
Db 1033 TGATCCACATCTACTGAG 1092

Qy 1286 ATGAAGACCAACCACTGCAAGCTGCTGCAATGTTGATGAGAGAGAGAGAGAGAGAG 1345  
Db 1093 ATGAAGACCAACCACTGCAAGCTGCTGCAATGTTGATGAGAGAGAGAGAGAGAGAG 1152  
Qy 1346 GAGCAAACTTAAAGTCTCCCTTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1405  
Db 1153 GAGCAAACTTAAAGTCTCCCTTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1212  
Qy 1406 GTCTGCCACCCGCTGAG 1465  
Db 1213 GTCTGCCACCCGCTGAG 1272  
Qy 1466 CAGAGAGTCTGAG 1525  
Db 1273 CAGAGAGTCTGAG 1332  
Qy 1526 TCCCTCCAAAG 1585  
Db 1333 TCCCTCCAAAG 1392  
Qy 1586 ACCGAGCTGAG 1645  
Db 1393 ACCGAGCTGAG 1452  
Qy 1646 GAAACAGAGTCAACCAAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1705  
Db 1453 GAAACAGAGTCAACCAAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1512  
Qy 1706 GGGCCCCGGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1765  
Db 1513 GGGCCCCGGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1572  
Qy 1766 GCTCTAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1825  
Db 1573 GCTCTAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1632  
Qy 1826 GTGGCTTCCGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1885  
Db 1633 GTGGCTTCCGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1692  
Qy 1886 GGCCTGGGGGGGCCCCCTGAGCCTTATGATGAGAGAGAGAGAGAGAGAGAGAGAG 1945  
Db 1693 GGCCTGGGGGGGCCCCCTGAGCCTTATGATGAGAGAGAGAGAGAGAGAGAGAGAG 1752  
Qy 1946 GAGAGCTGAG 2005  
Db 1753 GAGAGCTGAG 1812  
Qy 2006 AGATGCAAG 2055  
Db 1813 AGATGCAAG 1832

RESULT 8  
US-10-425-115-182496  
; Sequence 182496, Application US/10425115  
; Publication No. US20040214272A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa, Thomas J.  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Cao, Yongwei  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
; TITLE OF INVENTION: Plants  
; FILE REFERENCE: 38-21(5322)B  
; CURRENT APPLICATION NUMBER: US/10/425,115  
; NUMBER OF SEQ ID NOS: 2003-04-28  
; SEQ ID NO 182496  
; LENGTH: 763  
; TYPE: DNA  
; ORGANISM: Zea mays

FEATURE:  
NAME/KEY: unsure  
LOCATION: (1).(763)  
OTHER INFORMATION: unsure at all n locations  
FEATURE:  
OTHER INFORMATION: Clone ID: MRT4577\_98018C.1  
US-10-425-115-182496

Query Match 31.2% Score 631, DB 18; Length 763;  
Best Local Similarity 92.8% Pred. No. 9.3e-174;  
Matches 661; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

QY 29 GAGAGAAATGGGTCCAGGATTAAGTACCTATACCAAGCTGAGGACAGAGGCT 88  
DB 52 GCGAGAAATGGGTCCAGGATTAAGTACCTATACCAAGCTGAGGACAGAGGCT 111  
QY 89 ACAGTGTTCACCGCCGACCCCACTCAAGGATATGACAGACCCAGGATATGAGC 148  
DB 112 ACAGTGTTCACCGCCGACCCCACTCAAGGATATGACAGACCCAGGATATGAGC 171  
QY 149 AACAAAGCTATGAACTTATGACAGAGCCCTGATGTCAGCTTATCCAGGCTCAGACA 208  
DB 172 AACAAAGCTATGAACTTATGACAGAGCCCTGATGTCAGCTTATCCAGGCTCAGACA 231  
QY 209 CTGCAACCTATGGGACAGCCGCTATGACCTTCTTATGACAGCCCTCCAGCTGTTATA 268  
DB 232 CTGCAACCTATGGGACAGCCGCTATGACCTTCTTATGACAGCCCTCCAGCTGTTATA 291  
QY 269 CTACTCCAACTGCCCCCAGGACATACAGCCAGCTGTCAGGGATATGAGCTGTTCTT 328  
DB 292 GCACTCCAACTGCCCCCAGGACATACAGCCAGCTGTCAGGGATATGAGCTGTTCTT 351  
QY 329 ATGATACCACTGCTGCTACAGTACCAACCAAGGCTCTTATGACGCTCAGTCTGAT 388  
DB 352 ATGATACCACTGCTGCTACAGTACCAACCAAGGCTCTTATGACGCTCAGTCTGAT 411  
QY 389 ATGCACTGACGCTGCTTATGACGCTTATGAGGACAGGACCACTGCACTGCA 448  
DB 412 ATGCACTGACGCTGCTTATGACGCTTATGAGGACAGGACCACTGCACTGCA 471  
QY 449 GACCGAGATGAGAAACAGCCCACTGAGACTAGTCAACTCAATCTTATGACAGGGGTT 508  
DB 472 GACCGAGATGAGAAACAGCCCACTGAGACTAGTCAACTCAATCTTATGACAGGGGTT 531  
QY 509 ACAACCGCCGACCTGCTGATATGAGACAGGATATCACTATCCCAAGTACCTGGA 568  
DB 532 ATAAACCAACCGACCTGCTGATATGAGACAGGATATCACTATCCCAAGTACCTGGA 591  
QY 569 GCTAACCCATGACAGCAGTCACTGCACTTCACTCTTATCCCAAGTATCTCTTA 628  
DB 592 GCTAACCCATGACAGCAGTCACTGCACTTCACTCTTATCCCAAGTATCTCTTA 651  
QY 629 CACAGCCGACTGATATGATCAAGCACTTCTGACAGAAACCACTTATGGGCAACGA 688  
DB 652 CACAGCCGACTGATATGATCAAGCACTTCTGACAGAAACCACTTATGGGCAACGA 711  
QY 689 GCAGCTATGACAGCAAGTATGCTATGCAAAACCAAGTATGGGCAACA 740  
DB 712 GCAGCTATGACAGCAAGTATGCTATGCTATGCAAAACCAAGTATGGGCAACA 763

RESULT 9  
US-10-198-846-9847  
Sequence 9847, Application US/1019846  
Publication No. US2003009974A1  
GENERAL INFORMATION:  
APPLICANT: Lillie, James  
APPLICANT: Xu, Yongyao  
APPLICANT: Wang, Youzhen  
APPLICANT: Steinmann, Kathleen  
TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS  
FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND  
THERAPY OF BREAST CANCER

FILE REFERENCE: MRI-049  
CURRENT APPLICATION NUMBER: US/10/198,846  
CURRENT FILING DATE: 2002-07-18  
PRIOR APPLICATION NUMBER: 60/306,220  
PRIOR FILING DATE: 2001-07-18  
NUMBER OF SEQ ID NOS: 14084  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 9847  
LENGTH: 2299  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-10-198-846-9847

Query Match 28.1% Score 570; DB 14; Length 2299;  
Best Local Similarity 100.0% Pred. No. 9.9e-156;  
Matches 570; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1048 GAGCCATGATGAGAGACAGATCTTATCTAGGCCCACTGTATGATCAGATGAGAC 1107  
DB 1072 GAGCCATGATGAGAGACAGATCTTATCTAGGCCCACTGTATGATCAGATGAGAC 1131  
QY 1108 TGTGACACAGTGCATTTATGATCAAGGATTAATGACAGTGTGACTTATGATCTG 1167  
DB 1132 TGTGACACAGTGCATTTATGATCAAGGATTAATGACAGTGTGACTTATGATCTG 1191  
QY 1168 GCAAGCTTCTTAAAGAGTGGGGGTTTAAAGATGAAAGAGAACTGGGCAACCATG 1227  
DB 1192 GCAAGCTTCTTAAAGAGTGGGGGTTTAAAGATGAAAGAGAACTGGGCAACCATG 1251  
QY 1228 ATCCATCTTACCTGACAGAGAAACAGAAAGCCCAAGGCGATGCCAGTGTCTAT 1287  
DB 1252 ATCCATCTTACCTGACAGAGAAACAGAAAGCCCAAGGCGATGCCAGTGTCTAT 1311  
QY 1288 GAAAGCCCACTGCTGCAAGGCTGCTGGAATGTTGATGGGAAAGTTTCAAGG 1347  
DB 1312 GAAAGCCCACTGCTGCAAGGCTGCTGGAATGTTGATGGGAAAGTTTCAAGG 1371  
QY 1348 AGCAACTTAAAGTCTCCCTGCTGGAGAAAGCCCTCAATGAAAGTATGCGGGTGT 1407  
DB 1372 AGCAACTTAAAGTCTCCCTGCTGGAGAAAGCCCTCAATGAAAGTATGCGGGTGT 1431  
QY 1408 CTGCACTCCCTGAGAGGCAAGGCAATGCCACCACTCTGAGAGTCCAGAGGCCCA 1467  
DB 1432 CTGCACTCCCTGAGAGGCAAGGCAATGCCACCACTCTGAGAGTCCAGAGGCCCA 1491  
QY 1468 GAGAGTCTGAGAGACCAATGGGTGCAATGGAGGCGGTGAGAGATGAGAGGCTTC 1527  
DB 1492 GAGAGTCTGAGAGACCAATGGGTGCAATGGAGGCGGTGAGAGATGAGAGGCTTC 1551  
QY 1528 CTTCCAAAGAGACCCCGGGGTTCCGAGGGAACCCCTCTGAGAGAGAAACGTCCAGAC 1587  
DB 1552 CTTCCAAAGAGACCCCGGGGTTCCGAGGGAACCCCTCTGAGAGAGAAACGTCCAGAC 1611  
QY 1588 CGAGCTGAGAGCTGAGAGTGTCCCAATCCG 1617  
DB 1612 CGAGCTGAGAGCTGAGAGTGTCCCAATCCG 1641

RESULT 10  
US-09-864-761-8001  
Sequence 8001, Application US/09864761  
Patent No. US2002048763A1  
GENERAL INFORMATION:  
APPLICANT: Penn, Sharon G.  
APPLICANT: Rank, David R.  
APPLICANT: Hanzel, David K.  
APPLICANT: Chen, Wensheng  
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
FILE REFERENCE: Aeomica-X-1  
CURRENT APPLICATION NUMBER: US/09/864,761  
CURRENT FILING DATE: 2001-05-23  
PRIOR APPLICATION NUMBER: US 60/180,312



```

RESULT 11
US-09-918-995-27690
; Sequence 27690, Application US/09918995
; Publication No. US20030073623a1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; TITLE OF INVENTION: FROM VARIOUS CDNA LIBRARIES
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 27690
; LENGTH: 550
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(550)
; OTHER INFORMATION: n = A,T,C or G
US-09-918-995-27690

```

Query Match	25.1%	Score 508.4	DB 10	Length 550
Best Local Similarity	97.5%	Pred. No. 6e-138		
Matches 515	Conservative	0	Mismatches 13	Indels 0
			Gaps	0
QY	1327	GATGGAAAGATTTTCAGGGAGCAAACTTAAAGTCTCCCTTGCTCGGAAGAACCTCCA	1388	
Db	22	GAGAGGGCAATCTTTTCAGGGAGCAAACTTAAAGTCTCCCTTGCTCGGAAGAACCTCCA	81	
QY	1387	ATGAAACAGTATGCGGGGTGTCTGTCACCCCGTGAAGGGCAAGGCATGCCACCACTTC	1444	
Db	82	ATGAAACAGTATGCGGTGTGTCTGTCACCCCGTGAAGGGCAAGGCATGCCACCACTTC	141	
QY	1447	CGTGAAGGTCCAGAGAGGCCCAAGAGGTCTTGGGGGACCCATAGGTGTGCATTGGAGAGCCGT	1507	
Db	142	CGTGAAGGTCCAGAGAGGCCCAAGAGGTCTTGGGGGACCCATAGGTGTGCATTGGAGAGCCGT	201	
QY	1507	GGAGAGATPAGAGAGGAGCTTCCCTCCAGAGAGAACCCGGAGGCTTCCGAGGGAAACCCCTCT	1566	
Db	202	GGAGAGATPAGAGAGGAGCTTCCCTCCAGAGAGAACCCGGAGGCTTCCGAGGGAAACCCCTCT	261	
QY	1567	GGAGAGAGAAACGTCCAGCAACCGAGCTGGAGACTGGCAGTGTCCCAATCCGGGTGTGGGA	1628	
Db	262	GGAGAGAGAAACGTCCAGCAACCGAGCTGGAGACTGGCAGTGTCCCAATCCGGGTGTGGGA	321	
QY	1627	AACCAAACTTTCGCTCGAGAAACAAGTGCACCACTTAAAGGCCCAAGGCTTGAAGGC	1688	
Db	322	AACCAAACTTTCGCTCGAGAAACAAGTGCACCACTTAAAGGCCCAAGGCTTGAAGGC	381	

QY 1687 TTCTCTCCGCAACCCCTTTCGCCCGGGGTTGATCTGTGCAAGAGTGGCCCTGGTGGC 1746  
DB 382 TTCTCTCCGCAACCCCTTTCGCCCGGGGTTGATCTGTGCAAGAGTGGCCCTGGTGGC 441  
QY 1747 ATCGGGGAGGAGAGGTGGCTCATGATGTGTGTGTCCGGTGGATGTTCAAGAGT 1806  
DB 442 ATCGGGGAGGAGAGGTGGCTCATGATGTGTGTGTCCGGTGGATGTTCAAGAGT 501  
QY 1807 GGGCGGTGTGAGAGCAGAGGTGGCTTCGTGTGGCCGGGAGCATGAC 1854  
DB 502 GGGCGGTGTGAGAGCAGAGGTGGCTTCGTGTGGCCGGGAGCATGAC 549  
RESULT 12  
US-09-918-995-9557  
Sequence 9557, Application US/09918995  
Publication No. US20030073623A1  
GENERAL INFORMATION:  
APPLICANT: Hyseq, Inc.  
TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED  
FILE REFERENCE: 20411-756  
CURRENT APPLICATION NUMBER: US/09/918,995  
CURRENT FILING DATE: 2001-07-30  
PRIOR APPLICATION NUMBER: US/09/235,076  
PRIOR FILING DATE: 1999-01-20  
NUMBER OF SEQ ID NOS: 38054  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 9557  
LENGTH: 568  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: misc.feature  
LOCATION: (1) ... (568)  
OTHER INFORMATION: n = A, T, C or G  
US-09-918-995-9557  
Query Match 23.7%; Score 480.2; DB 10; Length 568;  
Best Local Similarity 95.5%; Pred. No. 1,1e-129;  
Matches 515; Conservative 0; Mismatches 20; Indels 4; Gaps 2;  
QY 932 ACCGGGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 991  
DB 33 ATCGGTGTGTGATTCGAGATTGATCTGTGAGGAGCATGAGAGGTGGCGAGAG 92  
QY 992 GAGGAGCGGTGTGATTCGAGATTGATCTGTGAGGAGGAGGAGGAGGAGGAGGAGGAG 1051  
DB 93 GAGGAGCGGTGTGATTCGAGATTGATCTGTGAGGAGGAGGAGGAGGAGGAGGAGGAG 149  
QY 1052 CCATGATGAGAGCAGATCTTATCTAGGCCCACTTATGATCCAGATGAGACTCTG 1111  
DB 150 CCATGATGAGAGCAGATCTTATCTAGGCCCACTTATGATCCAGATGAGACTCTG 209  
QY 1112 ACAACAGTGCAATTATGATCAAGATTTATGACGTGTGATCTTATGATCTTATGATCT 1171  
DB 210 ACAACAGTGCAATTATGATCAAGATTTATGACGTGTGATCTTATGATCTTATGATCT 269  
QY 1172 ACTTCTTAAAGATGTGGGTTGTTAAGATGAGCAAGAGGAGGAGGAGGAGGAGGAGGAG 1231  
DB 270 ACTTCTTAAAGATGTGGGTTGTTAAGATGAGCAAGAGGAGGAGGAGGAGGAGGAGGAG 329  
QY 1232 ACATCTAAGTGTGAGCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1291  
DB 330 ACATCTAAGTGTGAGCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 389  
QY 1292 ACCCAACCACTGCGCAAGGCTGCGTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1351  
DB 390 ACCCAACCACTGCGCAAGGCTGCGTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 449  
QY 1352 AACTTAAAGTCTCCCTTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1411

DB 450 AACTTAAAGTCTCCCTTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 508  
QY 1412 CACCCCGTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1470  
DB 509 CACCCCGTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 567  
RESULT 13  
US-09-864-761-7766/c  
Sequence 7766, Application US/09864761  
Patent No. US20020048763A1  
GENERAL INFORMATION:  
APPLICANT: Penn, Sharon G.  
APPLICANT: Rank, David R.  
APPLICANT: Hanzel, David K.  
APPLICANT: Chen, Wensheng  
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
FILE REFERENCE: Aecm1ca-X-1  
CURRENT APPLICATION NUMBER: US/09/864,761  
CURRENT FILING DATE: 2001-05-23  
PRIOR APPLICATION NUMBER: US 60/180,312  
PRIOR FILING DATE: 2000-02-04  
PRIOR APPLICATION NUMBER: US 60/207,456  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: US 09/632,366  
PRIOR FILING DATE: 2000-08-03  
PRIOR APPLICATION NUMBER: GB 24263.6  
PRIOR FILING DATE: 2000-10-04  
PRIOR APPLICATION NUMBER: US 60/236,359  
PRIOR FILING DATE: 2000-09-27  
PRIOR APPLICATION NUMBER: PCT/US01/00666  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00667  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00664  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00669  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00665  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00668  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00663  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00662  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00661  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00670  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: US 60/234,687  
PRIOR FILING DATE: 2000-09-21  
PRIOR APPLICATION NUMBER: US 09/608,408  
PRIOR FILING DATE: 2000-06-30  
PRIOR APPLICATION NUMBER: US 09/774,203  
PRIOR FILING DATE: 2001-01-29  
NUMBER OF SEQ ID NOS: 45117  
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1  
SEQ ID NO 7766  
LENGTH: 505  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
MAP TO AL162311.1  
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 4.5  
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2.5  
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2.6  
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.8  
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 3.4  
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 4.8  
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 4.7  
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.1

US-09-864-761-7766

Query Match 20.3%; Score 411.6; DB 9; Length 505;  
Best Local Similarity 90.3%; Pred. No. 1.2e-109;  
Matches 486; Conservative 0; Mismatches 19; Indels 33; Gaps 3;

QY 162 AACCTATGACAGCCCACTGATGTCACTATATACCAAGCTCAGACCACTGCAACCTTATG 221  
DB 505 AACCTATGACAGCCCACTGATGTCACTATATACCAAGCTCAGACCACTGCAACCTTATG 446  
QY 222 GCAGACCGCCTTATGCACTTCTTATGACAGCCTTCCCACTGGTTATATCTCACTC 281  
DB 445 G-----CTATGCACTTCTTATGACAGCCTTCCCACTGGTTATATCTCACTC 394  
QY 282 CCCCAGGCAATACAGCCAGCTGTCCAGGGGTATAGGCACTGTGCTTATATGATACCAAC 341  
DB 393 CCCCAGGCAATACAGCCAGCTGTCCAGGGGTATAGGCACTGTGCTTATATGATACCAAC 334  
QY 342 TGCTACAGTACCAACCAAGCCCTCTATGACAGTCACTGATGATATGAGCACTGAGCC 401  
DB 333 TGCTACAGTACCAACCAAGCCCTCTATGACAGTCACTGATGATATGAGCACTGAGCC 275  
QY 402 TGCTATTCAGCCTATGAGGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 461  
DB 274 TGCTATTCAGCCTATGAGGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 215  
QY 462 AAACAGCCCACTGAGACTGATCACTCACTCACTTATGACAGGGGGTTACACCAAGCCAG 521  
DB 214 TAACACACCCATGAGACTGATCACTCACTCACTTATGACAGGGGGTTACACCAAGCCAG 155  
QY 522 CCTAGATATGACAGAGTACTACTGATTTATCCAGGTTCTGGGAGCTACCCCATGCA 581  
DB 154 CTTATGATATGACAGAGTACTACTGATTTATCCAGGTTCTGGGAGCTACCCCATGCA 95  
QY 582 GCCAGTACTGCACTCTCACTCACTCTCTTACAGGATTTCTCTACAGAGCCGACTAG 641  
DB 94 GCCAGTACTGCACT 58

## RESULT 14

US-09-918-995-3868  
Sequence 3868, Application US/09918995  
Publication No. US20030073623A1  
GENERAL INFORMATION:  
APPLICANT: Hyseq, Inc.  
TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED  
FILE REFERENCE: FROM VARIOUS CDNA LIBRARIES  
CURRENT APPLICATION NUMBER: US/09/918,995  
CURRENT FILING DATE: 2001-07-30  
PRIOR APPLICATION NUMBER: US/09/235,076  
PRIOR FILING DATE: 1999-01-20  
NUMBER OF SEQ ID NOS: 38054  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 3868  
LENGTH: 401  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-09-918-995-3868

Query Match 19.7%; Score 399.4; DB 10; Length 401;  
Best Local Similarity 99.8%; Pred. No. 4.1e-106;  
Matches 400; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 25 GAAGGAGAGAAATGGCTGCTCCAGGATTAAGTACTTATAGCCAGAGTGCAGGCGAGAG 84  
DB 1 GAAGGAGAGAAATGGCTGCTCCAGGATTAAGTACTTATAGCCAGAGTGCAGGCGAGAG 60  
QY 85 GGCTACAGTGTCTTACACCGCCAGCCACTCAAGGATATGACAGACCAAGCCAGGATAT 144

## RESULT 15

US-09-864-761-24705  
Sequence 24705, Application US/09864761  
Patent No. US20020048763A1  
GENERAL INFORMATION:  
APPLICANT: Penn, Sharon G.  
APPLICANT: Rank, David R.  
APPLICANT: Hanzel, David K.  
APPLICANT: Chen, Wensheng  
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
FILE REFERENCE: Aeomica-X-1  
CURRENT APPLICATION NUMBER: US/09/864,761  
CURRENT FILING DATE: 2001-05-23  
PRIOR APPLICATION NUMBER: US 60/180,312  
PRIOR FILING DATE: 2000-02-04  
PRIOR APPLICATION NUMBER: US 60/207,456  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: US 09/632,366  
PRIOR FILING DATE: 2000-08-03  
PRIOR APPLICATION NUMBER: GB 24263,6  
PRIOR FILING DATE: 2000-10-04  
PRIOR APPLICATION NUMBER: US 60/236,359  
PRIOR FILING DATE: 2000-09-27  
PRIOR APPLICATION NUMBER: PCT/US01/00666  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00667  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00664  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00669  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00665  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00668  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00663  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00662  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00661  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00670  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: US 60/234,687  
PRIOR FILING DATE: 2000-09-21  
PRIOR APPLICATION NUMBER: US 09/608,408  
PRIOR FILING DATE: 2000-06-30  
PRIOR APPLICATION NUMBER: US 09/774,203

DB 61 GGCTACAGTGTCTTACACCGCCAGCCCACTCAAGATATGACAGACCACTCAGGACTAT 120  
QY 145 GGGCAACAAAGCTATGAACTTATGACAGCCCACTGATGTCACTTATACCAAGGCTCAG 204  
DB 121 GGGCAACAAAGCTATGAACTTATGACAGCCCACTGATGTCACTTATACCAAGGCTCAG 180  
QY 205 ACCACTGCAACTTATGAGCAGACCGCTATGCAACTTCTTATGACAGGCTTCCCACTGCT 264  
DB 181 ACCACTGCAACTTATGAGCAGACCGCTATGCAACTTCTTATGACAGGCTTCCCACTGCT 240  
QY 265 TATACTACTCACTGCCCCCAGGCAATACAGCAGCAGCTGTCCAGGGGTATGAGCACTGCT 324  
DB 241 TATACTACTCACTGCCCCCAGGCAATACAGCAGCAGCTGTCCAGGGGTATGAGCACTGCT 300  
QY 325 GCTTATGATACCAACCTGCTACAGTACCAACCAAGCCCTCTATGAGCACTGCT 384  
DB 301 GCTTATGATACCAACCTGCTACAGTACCAACCAAGCCCTCTATGAGCACTGCT 360  
QY 385 GCATATGGCACTGAGCTGCTTATCCAGCCTATGAGGCACTG 425  
DB 361 GCATATGGCACTGAGCTGCTTATCCAGCCTATGAGGCACTG 401

```
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annonmax Sequence Listing Engine vers. 1.1
; SEQ ID NO 24705
; LENGTH: 399
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC018774.3
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.1
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2.8
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 3.4
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 3
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 4.1
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2.4
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.9
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.6
; OTHER INFORMATION: SWISSPROT HIT: Q01844, EVALUE 3.00e-28
; OTHER INFORMATION: NT HIT: g11417927, EVALUE 0.00e+00
; OTHER INFORMATION: EST_HUMAN HIT: BE781742.1, EVALUE 0.00e+00
US-09-864-761-24705
```

```
Query Match 19.0%; Score 384.6; DB 9; Length 399;
Best Local Similarity 97.7%; Pred. No. 8.7e-102;
Matches 390; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 37 ATGGCGTCCAGGATTACAGTACTATAGCCAACTGAGGCGAGGCGTACAGTCT 96
DB 1 ATGGCATCCATGATTTACAGTACTATAGCCAACTGAGGCGAGGCGTACAGTCT 60
QY 97 TACACCGCCGAGCCCACTCAAGATATGACAGACCAACCAGGATATGGGCAACAAGC 156
DB 61 TACAACACCCAGCCCACTCAAGATATGACAGACCAACCAGGATATGGGCAACAAGC 120
QY 157 TATGAACTATGAGACAGCCCACTGATGTGCTATACCAAGGCTCAGACCACTGCAACC 216
DB 121 TATGAACTATGAGACAGCCCACTGATGTGCTATACCAAGGCTCAGACCACTGCAATC 180
QY 217 TATGGGCAAGCCGCTTATGCACTTCTTATGAGACAGGCTCCACTGGTTATACTGCA 276
DB 181 TATGGGCAAGCCGCTTATGCACTTCTTATGAGACAGGCTCCACTGGTTATACTGCA 240
QY 277 ACTGCCCCCGAGGATACAGCCAGCCTGTCCAGGGGTATGGCACTGGTGTATGATACC 336
DB 241 ACTGCCCCCGAGGATACAGCCAGCCTGTCCAGGGGTATGGCACTGGTGTATGATACC 300
QY 337 ACCACTGTACAGTACACCAACCAGGCTCTTATGAGGCTCAGTCTGCAATATGGCACT 396
DB 301 ACCACTGTACAGTACACCAACCAGGCTCTTATGAGGCTCAGTCTGCAATATGGCACT 360
QY 397 CAGCCTGCTTATCCAGGCTTATGGGCAAGCAGCAGCC 435
DB 361 CAGCCTGCTTATCCAGGCTTATGGGCAAGCAGCAGTAGCC 399
```

Search completed: February 20, 2005, 22:47:56  
Job time : 1071.67 secs



QY 61 TATAGCAAGCTGACGACGAGGCTACAGTCTTACACCGCCCAAGCCCACTCAAGGA 120  
DB 65 TATAGCAAGCTGACGACGAGGCTACAGTCTTACACCGCCCAAGCCCACTCAAGGA 124  
QY 121 TATGACAGACCAACCAAGGATATGAGCAACAAGCTATGAACTTATGACAGCCCACT 180  
DB 125 TATGACAGACCAACCAAGGATATGAGCAACAAGCTATGAACTTATGACAGCCCACT 184  
QY 181 GATGTCAGTATACCAAGCTCAGACCACTGCAACCTATGAGCAACCGCTATGCAACT 240  
DB 185 GATGTCAGTATACCAAGGCTCAGACCACTGCAACCTATGAGCAACCGCTATGCAACT 244  
QY 241 TCTTATGAGACGCTCCCACTGTTATACCTCACTCACTGCCCCCAAGGATACAGCAAG 300  
DB 245 TCTTATGAGACGCTCCCACTGTTATACCTCACTCACTGCCCCCAAGGATACAGCAAG 304  
QY 301 CCTGTCAGAGGATATGACACTGCTGCTTATGATACCACTGCTACAGTCAACCAACCC 360  
DB 305 CCTGTCAGAGGATATGACACTGCTGCTTATGATACCACTGCTACAGTCAACCAACCC 364  
QY 361 CAGGCTCTTATGACAGTCTGCTATGAGCACTGACCTGCTTATCCAGCTTATGAG 420  
DB 365 CAGGCTCTTATGACAGTCTGCTATGAGCACTGACCTGCTTATCCAGCTTATGAG 424  
QY 421 CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 480  
DB 425 CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 484  
QY 481 AGTCAACCTCAATCTAGACAGAGGAGTTACACCAAGCCAGCTAGGATATGACAGAGT 540  
DB 485 AGTCAACCTCAATCTAGACAGAGGAGTTACACCAAGCCAGCTAGGATATGACAGAGT 544  
QY 541 AACTCAAGTATATCCCAAGTACTGAGAGCTACCCATGACAGCCAGTCACTGACCTCCA 600  
DB 545 AACTCAAGTATATCCCAAGTACTGAGAGCTACCCATGACAGCCAGTCACTGACCTCCA 604  
QY 601 TCTTACCTCTCAACAGCTATCTCTCAACAGCCGCTAGTTATGATCAAGACAGTAC 660  
DB 605 TCTTACCTCTCAACAGCTATCTCTCAACAGCCGCTAGTTATGATCAAGACAGTAC 664  
QY 661 TCTCAGACGAACTATGAGGCAACAGAGAGCTATGACAGCAGTATGATGCTCA 720  
DB 665 TCTCAGACGAACTATGAGGCAACAGAGAGCTATGACAGCAGTATGATGCTCA 724  
QY 721 CAAAGCAGCTATGAGGCAACAGAGCTCCCACTAGTTACCAACCCAACTGATCTACAGC 780  
DB 725 CAAAGCAGCTATGAGGCAACAGAGCTCCCACTAGTTACCAACCCAACTGATCTACAGC 784  
QY 781 CAAAGCTCAAGTCAATATAGCAACAAGAGCAGCTTACGGGAGCAGAGTCACTTCCGA 840  
DB 785 CAAAGCTCAAGTCAATATAGCAACAAGAGCAGCTTACGGGAGCAGAGTCACTTCCGA 844  
QY 841 CAGGACCAACCCAGTACATGAGTGTATGAGGAGAGTCTGAGAGATTTTCCGACCA 900  
DB 845 CAGGACCAACCCAGTACATGAGTGTATGAGGAGAGTCTGAGAGATTTTCCGACCA 904  
QY 901 GAGAGAAACCGAGAGATGAGTGTCTGATTAACGGGGCAGGGGAAAGGGGGATTTGAT 960  
DB 905 GAGAGAAACCGAGAGATGAGTGTCTGATTAACGGGGCAGGGGAAAGGGGGATTTGAT 964  
QY 961 CGTGAAGGCAATGAGCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1020  
DB 965 CGTGAAGGCAATGAGCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1024  
QY 1021 GAGGAGAGTGGCTTCAATTAAGCTGTGAGCCCACTGATGAAGGACCAAGATCTTATGAT 1080  
DB 1025 GAGGAGAGTGGCTTCAATTAAGCTGTGAGCCCACTGATGAAGGACCAAGATCTTATGAT 1084  
QY 1081 GGGCCACCTGTGATCAAGATGAAGACTCTGACCAACAGTGAATTTATGTAACAAGATTA 1140  
DB 1085 GGGCCACCTGTGATCAAGATGAAGACTCTGACCAACAGTGAATTTATGTAACAAGATTA 1144

QY 1141 AATGACAGTGTGACTCTAGATGATCTGACAGACTTCTTTAAGCAGTGTGGGTGTTAAG 1200  
DB 1145 AATGACAGTGTGACTCTAGATGATCTGACAGACTTCTTTAAGCAGTGTGGGTGTTAAG 1204  
QY 1201 ATGAACAGAGAACTGGGCAACCATGATCCATCTTACCTTGAGCAAGAAACAGAAAG 1260  
DB 1205 ATGAACAGAGAACTGGGCAACCATGATCCATCTTACCTTGAGCAAGAAACAGAAAG 1264  
QY 1261 CCCAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1320  
DB 1265 CCCAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1324  
QY 1321 TGGTTTATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1380  
DB 1325 TGGTTTATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1384  
QY 1381 CCTCCAAATGAACATATGCGGGGAGTCTGCAACCCCTGAGAGGCAAGGATGCCAACA 1440  
DB 1385 CCTCCAAATGAACATATGCGGGGAGTCTGCAACCCCTGAGAGGCAAGGATGCCAACA 1444  
QY 1441 CCACTCCGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1500  
DB 1445 CCACTCCGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1504  
QY 1501 GGCCTGAGAGAGATGAGAGAGGCTTCCCTCAAGAGAGACCCCGGGGTTCCGAGGAGAC 1560  
DB 1505 GGCCTGAGAGAGATGAGAGAGGCTTCCCTCAAGAGAGACCCCGGGGTTCCGAGGAGAC 1564  
QY 1561 CCTCTGAG 1620  
DB 1565 CCTCTGAG 1624  
QY 1621 TGTGAGAAACAGAACTTGGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1680  
DB 1625 TGTGAGAAACAGAACTTGGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1684  
QY 1681 GAAGGCTTCTCCCGCACCTTCCCGCCCGGGGAGTGTATGAGTGGAGAGTGGCCCT 1740  
DB 1685 GAAGGCTTCTCCCGCACCTTCCCGCCCGGGGAGTGTATGAGTGGAGAGTGGCCCT 1744  
QY 1741 GGTGCAATCCGGGAG 1800  
DB 1745 GGTGCAATCCGGGAG 1804  
QY 1801 AGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1860  
DB 1805 AGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1864  
QY 1861 GGCCTTGTGAG 1920  
DB 1865 GGCCTTGTGAG 1924  
QY 1921 ATGAG 1980  
DB 1925 ATGAG 1984  
QY 1981 CAGGAGCGCAGAGATGAGGCTTACTAGATGACAGAGACCCCGCAGA 2025  
DB 1985 CAGGAGCGCAGAGATGAGGCTTACTAGATGACAGAGACCCCGCAGA 2029

RESULT 2  
LOCUS CR608375 2314 bp mRNA linear HTC 21-JUN-2004  
DEFINITION full-length cDNA clone CS0D1039J24 of Placenta Cot 25-normalized  
of Homo sapiens (human).  
ACCESSION CR608375  
VERSION CR608375.1 GI:50489182  
KEYWORDS HTC; CNSTL cDNA.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.







QY 1201 ATGAAACAAGAACTGGGCAACCCATGATTCACATCTACCTGAGCAAGAAACAGAGAAAG 1260  
DB 1207 ATGAAACAAGAACTGGGCAACCCATGATTCACATCTACCTGAGCAAGAAACAGAGAAAG 1266  
QY 1261 CCCAAGGCGATGCGACAGTGTCTCTATGAAAGCCACCCTGCAAGGCTGCGTGGAA 1320  
DB 1267 CCCAAGGCGATGCGACAGTGTCTCTATGAAAGCCACCCTGCAAGGCTGCGTGGAA 1326  
QY 1321 TGGTTTATGAGAAAGTTTTCAGAGGACAACTTAAAGTCTCTCTTCTGCGAAGAG 1380  
DB 1327 TGGTTTATGAGAAAGTTTTCAGAGGACAACTTAAAGTCTCTCTTCTGCGAAGAG 1386  
QY 1381 CCTCAATGAAACAGTATGCGGGGTGTCTGCAACCCGCTGAGGGGCAAGGATCCCA 1440  
DB 1387 CCTCAATGAAACAGTATGCGGGGTGTCTGCAACCCGCTGAGGGGCAAGGATCCCA 1446  
QY 1441 CCACTCCGTGAGAGTCCAGAGAGGCCAGAGGTCTGAGGAGCCCATGAGTCCGATGGA 1500  
DB 1447 CCACTCCGTGAGAGTCCAGAGAGGCCAGAGGTCTGAGGAGCCCATGAGTCCGATGGA 1506  
QY 1501 GGGCGTGAAGAGATTAAGAGAGGCTTCCCTCAAGAGAGCCCGGGGTTCGAGAGGAG 1560  
DB 1507 GGGCGTGAAGAGATTAAGAGAGGCTTCCCTCAAGAGAGCCCGGGGTTCGAGAGGAG 1566  
QY 1561 CCCCTGAGAGAGAGAAAGTCCAGAGAGGCCAGAGGTCTGAGGAGCCCATGAGTCCGATGGA 1620  
DB 1567 CCCCTGAGAGAGAGAAAGTCCAGAGAGGCCAGAGGTCTGAGGAGCCCATGAGTCCGATGGA 1626  
QY 1621 TGTGAAACCAAGAACTTCCCTGAGAGAGAGGTGCAACAGTGTAAAGGCCCAAGAGCT 1680  
DB 1627 TGTGAAACCAAGAACTTCCCTGAGAGAGAGGTGCAACAGTGTAAAGGCCCAAGAGCT 1686  
QY 1681 GAAGGCTTCTCCGCAACCTTTCGCGCCCGGGGTGTGATCTGTGCAAGAGTGGCT 1740  
DB 1687 GAAGGCTTCTCCGCAACCTTTCGCGCCCGGGGTGTGATCTGTGCAAGAGTGGCT 1746  
QY 1741 GGTGGCAATGCGGGGAGAGAAAGTGTGCTATGATGTGTGTCTCCGCTGAAATTTTC 1800  
DB 1747 GGTGGCAATGCGGGGAGAGAAAGTGTGCTATGATGTGTGTCTCCGCTGAAATTTTC 1806  
QY 1801 AGAGGTGCGCGTGTGAGACAGAGTGTGCTTCCGTGTGTGCGGGGATGAGCCAGGT 1860  
DB 1807 AGAGGTGCGCGTGTGAGACAGAGTGTGCTTCCGTGTGTGCGGGGATGAGCCAGGT 1866  
QY 1861 GGGCTTGTGTGAGAGAAAGAGTGTGCTTGTGAGGGGCGCCCTGTGACCTTTGATGAGACAG 1920  
DB 1867 GGGCTTGTGTGAGAGAAAGAGTGTGCTTGTGAGGGGCGCCCTGTGACCTTTGATGAGACAG 1926  
QY 1921 ATGAGAGAGAAAGAGAGAGCTGTGAGACCTGTGAAATGATTAAGGCGAGACCGT 1980  
DB 1927 ATGAGAGAGAAAGAGAGAGCTGTGAGACCTGTGAAATGATTAAGGCGAGACCGT 1986  
QY 1981 CAGAGAGCGAGAGTGGCCCTACTAGATGACAGAGCCCGCAGA 2025  
DB 1987 CAGAGAGCGAGAGTGGCCCTACTAGATGACAGAGCCCGCAGA 2031

RESULT 4  
LOCUS CR619493 2103 bp mRNA linear HTC 21-JUL-2004  
DEFINITION full-length cDNA clone CS0D1076Y021 of Placenta Cot 25-normalized  
of Homo sapiens (human).  
ACCESSION CR619493  
VERSION CR619493.1 GI:50500300  
KEYWORDS HTC; CNSLT cDNA.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 2103)  
AUTHORS Li, W.B., Gruber, C., Jessee, J. and Polayes, D.  
TITLE Full-length cDNA libraries and normalization

JOURNAL Unpublished  
REMARK Contact : Feng Liang Email : fliang@lifetech.com URL :  
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600  
REFERENCE Faraday Avenue  
AUTHORS 2 (bases 1 to 2103)  
TITLE Genoscope.  
JOURNAL Direct Submission  
Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage :  
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr  
- Web : www.genoscope.cns.fr)  
COMMENT 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime  
end enriched, double-strand cDNA was digested with Not I and cloned  
into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library  
was normalized. Library was constructed by Life Technologies, a  
division of Invitrogen.  
FEATURES  
Location/Qualifiers  
source 1..2103  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="CS0D1076Y021"  
/issue\_type="Placenta Cot 25-normalized"  
/plasmid="pCMVSPORT\_6"

ORIGIN  
Query Match 97.6%; Score 1977; DB 3; Length 2103;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1977; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 49 GATTACAGTACCTATGCGAGCGGACAGGCGCTACAGTGTTCACCGCCAG 108  
DB 1 GATTACAGTACCTATGCGAGCGGACAGGCGCTACAGTGTTCACCGCCAG 60  
QY 109 CCCACTCAGATATGACAGACCAACCAAGCAATATGAGCAAAAGCTATGAACTTAT 168  
DB 61 CCCACTCAGATATGACAGACCAACCAAGCAATATGAGCAAAAGCTATGAACTTAT 120  
QY 169 GAGACGCCATGATGTGAGCTATACCAAGGCTTACACCTGCAACTTATGAGGAGACC 228  
DB 121 GAGACGCCATGATGTGAGCTATACCAAGGCTTACACCTGCAACTTATGAGGAGACC 180  
QY 229 GCGTATGCACTTCTATGAGAGCGCTCCACGTGTATCTCAATGCGCCCGCAG 288  
DB 181 GCGTATGCACTTCTATGAGAGCGCTCCACGTGTATCTCAATGCGCCCGCAG 240  
QY 289 GCATTAAGAGCGCTGTCCAGGGGTATGAGCACTGTGCTTATATGACCACTGCTACA 348  
DB 241 GCATTAAGAGCGCTGTCCAGGGGTATGAGCACTGTGCTTATATGACCACTGCTACA 300  
QY 349 GTACCAACCAACCGGCTCTTATGAGAGCTCACTGTGATATGAGCACTGAGCTTAT 408  
DB 301 GTACCAACCAACCGGCTCTTATGAGAGCTCACTGTGATATGAGCACTGAGCTTAT 360  
QY 409 CCAAGCTTATGAG 468  
DB 361 CCAAGCTTATGAG 420  
QY 469 CCAAGCTTATGAG 528  
DB 421 CCAAGCTTATGAG 480  
QY 529 TATGACAGAGTACAGTATATCCCAAGTATCCCAAGTATCCCAAGTATCCCAAGTAT 588  
DB 481 TATGACAGAGTACAGTATATCCCAAGTATCCCAAGTATCCCAAGTATCCCAAGTAT 540  
QY 589 ACTGACCTTCATCTCACTCTCTTACAGAGTATCTCTTCAACAGCGGATATGATGAT 648  
DB 541 ACTGACCTTCATCTCACTCTCTTACAGAGTATCTCTTCAACAGCGGATATGATGAT 600  
QY 649 CAGAGAGTTACTCTGAGAGAAACCTTATGAGGCAACGAGAGCTATGAGAGAGAGT 708  
DB 601 CAGAGAGTTACTCTGAGAGAAACCTTATGAGGCAACGAGAGCTATGAGAGAGAGT 660

QY 709 AGCTATGTCACCAAGACAGCTATGAGGAGCAGCTCTCCACTGATTACCAACCCCAACT 768  
 DB 661 AGCTATGTCACCAAGACAGCTATGAGGAGCAGCTCTCCACTGATTACCAACCCCAACT 720  
 QY 769 GGTATCTACAGCCAGCTCTCAAGTCAATATAGCCAAAGAGCAGCAGCTACCGGAGCAG 828  
 DB 721 GGTATCTACAGCCAGCTCTCAAGTCAATATAGCCAAAGAGCAGCAGCTACCGGAGCAG 780  
 QY 829 AGTTCAATCCGACAGAGCAGCCCACTGATGATGAGGAGGAGTCTGAGAG 888  
 DB 781 AGTTCAATCCGACAGAGCAGCCCACTGATGATGAGGAGGAGTCTGAGAG 840  
 QY 889 TTTTCGAGACAGAGAGAGAGCCGAGAGATGATGAGGAGGAGTCTGAGAG 948  
 DB 841 TTTTCGAGACAGAGAGAGAGCCGAGAGATGATGAGGAGGAGTCTGAGAG 900  
 QY 949 GGGGGATTTGATCTGAGAGGAGATGAGAGAGTGGGAGGAGAGAGAGAGAGAGAG 1008  
 DB 901 GGGGGATTTGATCTGAGAGGAGATGAGAGAGTGGGAGGAGAGAGAGAGAGAGAG 960  
 QY 1009 GGCAGAGCTGAGAGAGAGAGTGGCTTCAATATAGCCCAAGCTGATGAGAGAGAG 1068  
 DB 961 GGCAGAGCTGAGAGAGAGAGTGGCTTCAATATAGCCCAAGCTGATGAGAGAGAG 1020  
 QY 1069 GATCTTATGATGAGAGAGAGAGTGGCTTCAATATAGCCCAAGCTGATGAGAGAGAG 1128  
 DB 1021 GATCTTATGATGAGAGAGAGAGTGGCTTCAATATAGCCCAAGCTGATGAGAGAGAG 1080  
 QY 1129 GTACAGAGATTAATGACAGAGAGAGTGGCTTCAATATAGCCCAAGCTGATGAGAGAG 1188  
 DB 1081 GTACAGAGATTAATGACAGAGAGAGTGGCTTCAATATAGCCCAAGCTGATGAGAGAG 1140  
 QY 1189 GGGGTTGTTAAGTGAAG 1248  
 DB 1141 GGGGTTGTTAAGTGAAG 1200  
 QY 1249 GAAACAG 1308  
 DB 1201 GAAACAG 1260  
 QY 1309 GCTGCGGTGAG 1368  
 DB 1261 GCTGCGGTGAG 1320  
 QY 1369 GCTGCGGTGAG 1428  
 DB 1321 GCTGCGGTGAG 1380  
 QY 1429 GGCATGAC 1488  
 DB 1381 GGCATGAC 1440  
 QY 1489 GGTGCGATGAG 1548  
 DB 1441 GGTGCGATGAG 1500  
 QY 1549 TCCGAG 1608  
 DB 1501 TCCGAG 1560  
 QY 1609 CCCAATCCGAGGTTGAG 1668  
 DB 1561 CCCAATCCGAGGTTGAG 1620  
 QY 1669 GCCCAAG 1728  
 DB 1621 GCCCAAG 1680  
 QY 1729 AGAGGTGAG 1788  
 DB 1681 AGAGGTGAG 1740  
 QY 1789 GGTGAGATGTTCAAG 1848

DB 1741 GGTGAGATGTTCAAG 1800  
 QY 1849 ATGAGACAG 1908  
 DB 1801 ATGAGACAG 1860  
 QY 1909 TTGATGAG 1968  
 DB 1861 TTGATGAG 1920  
 QY 1969 GGCAGAGACAG 2028  
 DB 1921 GGCAGAGACAG 1977  
 RESULT 5  
 CR602561  
 LOCUS  
 DEFINITION  
 full-length cDNA clone CS0D1068YN23 of Placenta Cot 25-normalized  
 of Homo sapiens (human).  
 ACCESSION  
 CR602561  
 VERSION  
 CR602561.1 GI:50483368  
 KEYWORDS  
 HTC; CNSLT\_cDNA.  
 SOURCE  
 Homo sapiens (human)  
 ORGANISM  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE  
 1 (bases 1 to 2070)  
 Li, W.B., Gruber, C., Jesse, J. and Polayes, D.  
 Full-length cDNA libraries and normalization  
 Unpublished  
 JOURNAL  
 Contact : Feng Liang Email : fliang@life.techn.com URL :  
 http://fulllength.invitrogen.com/ Invitrogen Corporation 1600  
 Faraday Avenue  
 Redwood City, CA 94063  
 REFERENCE  
 2 (bases 1 to 2070)  
 Genoscope.  
 Direct Submission  
 Submitted (20-JUN-2004) Genoscope - Centre National de Sequencage :  
 JOURNAL  
 BP 191 91006 Evry cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)  
 COMMENT  
 - Web : www.genoscope.cns.fr  
 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime  
 end enriched, double-strand cDNA was digested with Not I and cloned  
 into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library  
 was normalized. Library was constructed by Life Technologies, a  
 division of Invitrogen.  
 FEATURES  
 Location/Qualifiers  
 source  
 1..2070  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="CS0D1068YN23"  
 /cissue\_type="Placenta Cot 25-normalized"  
 /plasmid="pCMVSPORT\_6"  
 ORIGIN  
 Query Match 96.8%; Score 1961; DB 3; Length 2070;  
 Best Local Similarity 99.8%; Pred. No. 0;  
 Matches 1974; Conservative 0; Mismatches 0; Indels 3; Gaps 1;  
 QY 49 GATTACAGTACCTATGAG 108  
 DB 1 GATTACAGTACCTATGAG 60  
 QY 109 CCCACTCAAGAGATGACAG 168  
 DB 61 CCCACTCAAGAGATGACAG 120  
 QY 169 GAG 228  
 DB 121 GAG 180  
 QY 229 GCTATGCAACTTCTATGAG 288

Db 181 GCGTATGCACTTCTTATGAGCAGCTCCCACTGGTTATACCTACCTAGCCCCAG 240  
Qy 289 GCATACAGCCAGCCTGTCCAGGGGTATGCACTGGTCTTATGATACCACTGTACA 348  
Db 241 GCATACAGCCAGCCTGTCCAGGGGTATGCACTGGTCTTATGATACCACTGTACA 300  
Qy 349 GTACACCAACCCAGGCTCTTATGCACTGTCTGATATGCACTGACCTGTAT 408  
Db 301 GTACACCAACCCAGGCTCTTATGCACTGTCTGATATGCACTGACCTGTAT 360  
Qy 409 CCAGCCTATGGGAGCAGCCAGCAGCCATGCAAGACCGCAGATGAGAAAG 468  
Db 361 CCAGCCTATGGGAGCAGCCAGCAGCCATGCAAGACCGCAGATGAGAAAG 420  
Qy 469 CCCACTGAGACTAGTCAACTCAATCTGAGCAGGGGGTTACAACAGCCAGCTTGA 528  
Db 421 CCCACTGAGACTAGTCAACTCAATCTGAGCAGGGGGTTACAACAGCCAGCTTGA 480  
Qy 529 TATGACAGATTAATCACTATCCCAAGTACTGGAGCTACCCATGCAAGCCAGTC 588  
Db 481 TATGACAGATTAATCACTATCCCAAGTACTGGAGCTACCCATGCAAGCCAGTC 540  
Qy 589 ACTGACCTTCATCTTACCTCTCAACAGGCTATTCCTCTACAGAGCCGACTAGTTATGAT 648  
Db 541 ACTGACCTTCATCTTACCTCTCAACAGGCTATTCCTCTACAGAGCCGACTAGTTATGAT 600  
Qy 649 CAGAGCAGTTACTCTCAGCAAGAACCTATGGGCAACGAGCAGCTATGCAAGCAGAT 708  
Db 601 CAGAGCAGTTACTCTCAGCAAGAACCTATGGGCAACGAGCAGCTATGCAAGCAGAT 660  
Qy 709 AGCTATGCTCAACCAAGCACTATGGGCAAGCCTCCCACTAGTTACCAACCCAACT 768  
Db 661 AGCTATGCTCAACCAAGCACTATGGGCAAGCCTCCCACTAGTTACCAACCCAACT 720  
Qy 769 GGATCCCTACAGCAAGCTCAATATAGCAAGAGCAGCAGCTACGGGAGCAG 828  
Db 721 GGATCCCTACAGCAAGCTCAATATAGCAAGAGCAGCAGCTACGGGAGCAG 780  
Qy 829 AGTTCAATCCGAGAGCAACCCAGTATGAGTGGTGTATATGGCAGAGTCTGAGGA 888  
Db 781 AGTTCAATCCGAGAGCAACCCAGTATGAGTGGTGTATATGGCAGAGTCTGAGGA 840  
Qy 889 TTTTCCGAGCAGAGAGAACCCGAGTATGAGTGGTGTATATGGCAGAGGAGGA 948  
Db 841 TTTTCCGAGCAGAGAGAACCCGAGTATGAGTGGTGTATATGGCAGAGGAGGA 900  
Qy 949 GGGGATTTGATGTGGAGCATGAGCAGGTGGGCGGGAGAGAGAGCGGGTGAATG 1008  
Db 901 GGGGATTTGATGTGGAGCATGAGCAGGTGGGCGGGAGAGAGAGCGGGTGAATG 960  
Qy 1009 GGGAGCGCTGAGAGCGAGGTGGCTTCAATATAGCCTGGTGAACCCATGATGAGAGCA 1068  
Db 961 G---GGGCTGGAAGAGAGGTGGCTTCAATATAGCCTGGTGAACCCATGAGAGCA 1017  
Qy 1069 GATCTTGAATTAGCCCACTGTATGATCCAGTGAAGACTCTGCAACAGTGAATTTAT 1128  
Db 1018 GATCTTGAATTAGCCCACTGTATGATCCAGTGAAGACTCTGCAACAGTGAATTTAT 1077  
Qy 1129 GTACAGATTAATATGCACTGTGATCTGATGATCTGCAACTCTTATAGCAGTGT 1188  
Db 1078 GTACAGATTAATATGCACTGTGATCTGATGATCTGCAACTCTTATAGCAGTGT 1137  
Qy 1189 GGGGTTGTTAAGATGAGCAAGAGAACTGGCAACCCATGATCACTACTGAGCAAG 1248  
Db 1138 GGGGTTGTTAAGATGAGCAAGAGAACTGGCAACCCATGATCACTACTGAGCAAG 1197  
Qy 1249 GAAACAGAGAAAGCCCAAGGCGATGCCAGTGTCTTATGAAAGCCCACTGCCAG 1308  
Db 1198 GAAACAGAGAAAGCCCAAGGCGATGCCAGTGTCTTATGAAAGCCCACTGCCAG 1257  
Qy 1309 GCTGCGTGAATGTTTATGAGAAAGATTTTCAAGGGAGCAACTTAAGTCTCCCTT 1368

Db 1258 GCTGCCGTGAATGCTTTGATGGGAAAGATTTTCAAGGAGCAAACTTAAGTCTCCCTT 1317  
Qy 1369 GCTCGAAGAGAGCTTCCAAATGAAACAGTATGCGGGGTGTCTGCCAACCCGTGAGGAGGA 1428  
Db 1318 GCTCGAAGAGAGCTTCCAAATGAAACAGTATGCGGGGTGTCTGCCAACCCGTGAGGAGGA 1377  
Qy 1429 GGCATGCCAACCAACTCCCTGAGAGTCCAGAGAGCCCAAGAGATCCTGGGGAGACCCATG 1488  
Db 1378 GGCATGCCAACCAACTCCCTGAGAGTCCAGAGAGCCCAAGAGATCCTGGGGAGACCCATG 1437  
Qy 1489 GGTGCAATGGGAGGCGGTGAGAGATGAGAGAGCTTCCCTCAAGAGAGACCCGGGGT 1548  
Db 1438 GGTGCAATGGGAGGCGGTGAGAGATGAGAGAGCTTCCCTCAAGAGAGACCCGGGGT 1497  
Qy 1549 TCCGAGAGGAAACCTCTGAGAGAGAAAGTCCAGACCCGAGCTGAGACTGAGTGT 1608  
Db 1498 TCCGAGAGGAAACCTCTGAGAGAGAAAGTCCAGACCCGAGCTGAGACTGAGTGT 1557  
Qy 1609 CCCAATCCGGGTGTGAGAAACAGAACTTGTGCTGAGAGAAACAGATGCAACAGTATAG 1668  
Db 1558 CCCAATCCGGGTGTGAGAAACAGAACTTGTGCTGAGAGAAACAGATGCAACAGTATAG 1617  
Qy 1669 GCCCAAGAGCTGAGAGGCTTCTCCGCGCAACCTTTCGCGCCCGGGTGTGATCGTGGC 1728  
Db 1618 GCCCAAGAGCTGAGAGGCTTCTCTCCGCGCAACCTTTCGCGCCCGGGTGTGATCGTGGC 1677  
Qy 1729 AGAGTGGCCCTGTGTGATGCGGGAGAGAAAGTGGCTCTATGATCGTGTGTCCC 1788  
Db 1678 AGAGTGGCCCTGTGTGATGCGGGAGAGAAAGTGGCTCTATGATCGTGTGTCCC 1737  
Qy 1789 GGTGAAATGTTCAAGAGTGGCCGTGTGAGAACAGAGTGGCTTCCGTGTGTGGCCGGGGC 1848  
Db 1738 GGTGAAATGTTCAAGAGTGGCCGTGTGAGAACAGAGTGGCTTCCGTGTGTGGCCGGGGC 1797  
Qy 1849 ATGACCCAGAGTGGCTTGTGTGAGAGAAAGCAGAGTGGCCCTGTGGGGGCCCTGTGACT 1908  
Db 1798 ATGACCCAGAGTGGCTTGTGTGAGAGAAAGCAGAGTGGCCCTGTGGGGGCCCTGTGACT 1857  
Qy 1909 TTGATGGAACAGATGGAGAGAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGATGAT 1968  
Db 1858 TTGATGGAACAGATGGAGAGAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGATGAT 1917  
Qy 1969 GGGGACACCGGTGAG 2025  
Db 1918 GGGGACACCGGTGAG 1974

RESULT 6  
CR594868  
LOCUS 2093 bp mRNA linear HTC 21-JUL-2004  
DEFINITION Full-length cDNA clone CS0D1064YE15 of Placenta Cot 25-normalized  
of Homo sapiens (human).  
ACCESSION CR594868  
VERSION CR594868.1 GI:50475675  
KEYWORDS HTC; CNSLT\_cDNA.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE  
AUTHORS Li, W.B., Gruber, C., Jessee, J., and Polyes, D.  
TITLE Full-length cDNA libraries and normalization  
JOURNAL Unpublished  
REMARK Contact : Feng Liang Email : fliang@lifetech.com URL :  
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600  
Faraday Avenue  
2 (bases 1 to 2093)  
REFERENCE  
AUTHORS Genoscope.  
TITLE Direct Submision  
JOURNAL Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage :  
BP 191 91006 EVRY cedex - FRANCE [E-mail : seqref@genoscope.cns.fr  
- Web : www.genoscope.cns.fr]  
COMMENT 1st strand cDNA was primed with a Nott-oligo(dT) primer. Five prime



end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies, a division of Invitrogen.

FEATURES  
source  
Location/Qualifiers  
1..2093

/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="CS01064YE15"  
/tissue\_type="Placenta Cot 25-normalized"  
/plasmid="pCMVSPORT\_6"

## ORIGIN

Query Match 96.8%; Score 1961; DB 3; Length 2093;  
Best Local Similarity 99.8%; Pred. No. 0;  
Matches 1974; Conservative 0; Mismatches 0; Indels 3; Gaps 1;

49 GATTACAGTACCTATGCGAAGCTGAGCGGAGGAGGCTACAGTGTCTTACACCGCCAG 108  
1 GATTACAGTACCTATGCGAAGCTGAGCGGAGGAGGCTACAGTGTCTTACACCGCCAG 60  
109 CCCACTCAAGATATGACAGACACCCAGGCAATATGGCAAAAGCTATGGAACTTAT 168  
61 CCCACTCAAGATATGACAGACACCCAGGCAATATGGCAAAAGCTATGGAACTTAT 120  
169 GGAACAGCCCACTGATGCTATACCCAGGCTCAAGACCTGCAACCTATGGGAGACC 228  
121 GGAACAGCCCACTGATGCTATACCCAGGCTCAAGACCTGCAACCTATGGGAGACC 180  
229 GCGTATGCAACTCTTATGAGCAGCCCTCCAGTGTATTAATCTCAACTGCCCCAG 288  
181 GCGTATGCAACTCTTATGAGCAGCCCTCCAGTGTATTAATCTCAACTGCCCCAG 240  
289 GCATACAGCCAGCCTGTCAGGGGTATGGCATGTGTCTTATGATACCACTGCTTAC 348  
241 GCATACAGCCAGCCTGTCAGGGGTATGGCATGTGTCTTATGATACCACTGCTTAC 300  
349 GTACACACCAACCCAGGCTCTTATGAGCAGCTTGTGCAATATGGCACTGCTTAT 408  
301 GTACACACCAACCCAGGCTCTTATGAGCAGCTTGTGCAATATGGCACTGCTTAT 360  
409 CCAGCCTATGGGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 468  
361 CCAGCCTATGGGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 420  
469 CCCACTGAGACTAGTCAACTCTTATGAGCAGGGGTATCAACCGACCCAGCTTAT 528  
421 CCCACTGAGACTAGTCAACTCTTATGAGCAGGGGTATCAACCGACCCAGCTTAT 480  
529 TATGAGACAGTAACTTATGAGTATCCCAAGTACCTGGGAGCTAACCCCAAGCAGATC 588  
481 TATGAGACAGTAACTTATGAGTATCCCAAGTACCTGGGAGCTAACCCCAAGCAGATC 540  
589 ACTGACCTCATCTCTACCTCTTACAGAGCTATCTCTTACAGAGCCGCTAGTTATGAT 648  
541 ACTGACCTCATCTCTACCTCTTACAGAGCTATCTCTTACAGAGCCGCTAGTTATGAT 600  
649 CAGAGCAGTAACTCTTACAGAGAACTCTTATGGGCAACGAGCAGTATGGACAGCAGAT 708  
601 CAGAGCAGTAACTCTTACAGAGAACTCTTATGGGCAACGAGCAGTATGGACAGCAGAT 660  
709 AGCTATGCTGCAAAAGCAGTATGGGCAACGAGCAGTATGGGCAACGAGCAGTATGG 768  
661 AGCTATGCTGCAAAAGCAGTATGGGCAACGAGCAGTATGGGCAACGAGCAGTATGG 720  
769 GATTCCTACAGCAGCAGTCAAGTCAATATAGCAACAGAGCAGCAGCAGCAGCAGCAG 828  
721 GATTCCTACAGCAGCAGTCAAGTCAATATAGCAACAGAGCAGCAGCAGCAGCAGCAG 780  
829 AGTTCATTCGAGACAGCAGCAGCAGTATGAGTGTATTTATGGCAGAGTCTGAGGA 888  
781 AGTTCATTCGAGACAGCAGCAGCAGTATGAGTGTATTTATGGCAGAGTCTGAGGA 840

889 TTTTCGAGCAGAGAGAGACCGAGCATGATGGCCCTGATTAACCGGGCAGGGGAGA 948  
841 TTTTCGAGCAGAGAGAGACCGAGCATGATGGCCCTGATTAACCGGGCAGGGGAGA 900  
949 GGGGAGATTGATCTGAGAGCAGTATGAGCAGAGGTGGGCGGGAGAGAGACCGGCTGAATG 1008  
901 GGGGAGATTGATCTGAGAGCAGTATGAGCAGAGGTGGGCGGGAGAGAGACCGGCTGAATG 960  
1009 GGCAGCGCTGAGAGAGCAGAGTGGCTTCAATAAGCTCTGTGAGACCTGATTAAGAGCA 1068  
961 G--GCGCTGAGAGCAGAGTGGCTTCAATAAGCTCTGTGAGACCTGATTAAGAGCA 1017  
1069 GATCTTATCTAGGCGCCACCTGATAGTCCAGATGAAGACTCTGACCAAGTGAATTTAT 1128  
1018 GATCTTATCTAGGCGCCACCTGATAGTCCAGATGAAGACTCTGACCAAGTGAATTTAT 1077  
1129 GTACAAAGATTAAATGACAGTGTGACTTATGATCTGAGCAGACTTCTTTAAGCAGTGT 1188  
1078 GTACAAAGATTAAATGACAGTGTGACTTATGATCTGAGCAGACTTCTTTAAGCAGTGT 1137  
1189 GGGGTGTTAATGATGAACAAGAACTGGGCAACCAATGATCAATCTTACCTGAGCAAG 1248  
1138 GGGGTGTTAATGATGAACAAGAACTGGGCAACCAATGATCAATCTTACCTGAGCAAG 1197  
1249 GAAACAGAAAGCCCAAGAGCAGATGTCCTATGAAGACCAACCACTGACCAAG 1308  
1198 GAAACAGAAAGCCCAAGAGCAGATGTCCTATGAAGACCAACCACTGACCAAG 1257  
1309 GCTGCGGTGAATGCTTATGAGGAAAGATTTTCAAGGAGCAAACTTAAAGTCTCCCTT 1368  
1258 GCTGCGGTGAATGCTTATGAGGAAAGATTTTCAAGGAGCAAACTTAAAGTCTCCCTT 1317  
1369 GCTCGGAAAGACCTTCCCAATGAACAGTATCGGGGTGTCTGCAACCCGTGAGGGCAAG 1428  
1318 GCTCGGAAAGACCTTCCCAATGAACAGTATCGGGGTGTCTGCAACCCGTGAGGGCAAG 1377  
1429 GGCATGCAACCAACTCCGTGAGAGGTCCAGAGGCCAGAGAGTCTTGGGAGACCCATG 1488  
1378 GGCATGCAACCAACTCCGTGAGAGGTCCAGAGGCCAGAGAGTCTTGGGAGACCCATG 1437  
1489 GGTGCGATGGAGGCCGTGAGAGATGAGAGAGCTTCTTCAAGAGAGACCCCGGGGT 1548  
1438 GGTGCGATGGAGGCCGTGAGAGATGAGAGAGCTTCTTCAAGAGAGACCCCGGGGT 1497  
1549 TCCGAGGGAACCCCTCTGAGAGAGAAAGTCCAGACCCGAGCTGAGAGCTGGCAGTGT 1608  
1498 TCCGAGGGAACCCCTCTGAGAGAGAAAGTCCAGACCCGAGCTGAGAGCTGGCAGTGT 1557  
1609 CCCAATCCGGGTTGTGAAACAGAACTTCCGCTGAGAGACAGATGCAACAGTGTAA 1668  
1558 CCCAATCCGGGTTGTGAAACAGAACTTCCGCTGAGAGACAGATGCAACAGTGTAA 1617  
1669 GCCCAAAAGCTGAAGGCTTCTCCGCGACCTTTCCGCGCCCGGGGTGATCTGTGAGC 1728  
1618 GCCCAAAAGCTGAAGGCTTCTCCGCGACCTTTCCGCGCCCGGGGTGATCTGTGAGC 1677  
1729 AAGAGTGGCCCTGTGTGAGTATGCGGGAGAGAAAGTGGCTCAATGATCTGTGTGGTCCC 1788  
1678 AAGAGTGGCCCTGTGTGAGTATGCGGGAGAGAAAGTGGCTCAATGATCTGTGTGGTCCC 1737  
1789 GGTGAATGTTTCAAGAGTGGCCGTGTGAGACAGAGTGGCTTCCGTGTGTGAGCCGGGCG 1848  
1738 GGTGAATGTTTCAAGAGTGGCCGTGTGAGACAGAGTGGCTTCCGTGTGTGAGCCGGGCG 1797  
1849 ATGACCGAGGTGGCTTGTGTGAGAGAGAGAGTGGCTTGTGGGGGCGCCCTGTGAGCT 1908  
1798 ATGACCGAGGTGGCTTGTGTGAGAGAGAGAGTGGCTTGTGGGGGCGCCCTGTGAGCT 1857  
1909 TTGATGGAACAGATGGAG 1968  
1858 TTGATGGAACAGATGGAG 1917







Db 961 GTGAGCCCATGATGAAAGACAGATCTTGATAGGCCACCTGTGATCCAGATGAAG 1020  
 QY 1106 ACTGTGACAACTGATGAAATTTATGTAACAAGATTTAAATGACATGTGACTCTGATGATC 1165  
 Db 1021 ACTGTACAAAGTGTGAAATTTATGTAACAAGATTTAAATGACATGTGACTCTGATGATC 1080  
 QY 1166 TGGCAGACTCTTTAAAGCAGTGTGGGGTGTGTAAGATGAAACAAGAACTGGGCAACCC 1225  
 Db 1081 TGGCAGACTCTTTAAAGCAGTGTGGGGTGTGTAAGATGAAACAAGAACTGGGCAACCC 1140  
 QY 1226 TGATCCACTTACTGTAAGCAAGAAACAGAAAGCCCAAGGCGATGCCACAGTGTCT 1285  
 Db 1141 TGATCCACTTACTGTAAGCAAGAAACAGAAAGCCCAAGGCGATGCCACAGTGTCT 1200  
 QY 1286 ATGAAAGCCCAACCCACTGCGCAAGGCTGCGCTGTGAAATGTTGATGGGAAAGATTTTCAAG 1345  
 Db 1201 ATGAAAGCCCAACCCACTGCGCAAGGCTGCGCTGTGAAATGTTGATGGGAAAGATTTTCAAG 1260  
 QY 1346 GGAGCAAACTTAAAGTCTCCCTTGTCTGGAAGAAAGCTCCAAATGAAACAAGTATGCGGGGTG 1405  
 Db 1261 GGAGCAAACTTAAAGTCTCCCTTGTCTGGAAGAAAGCTCCAAATGAAACAAGTATGCGGGGTG 1320  
 QY 1406 GTCTGCAACCCCGTGAAGGCGAAGGCAATGCCACCACTCCGTGAAGGTCCAGAGGCC 1465  
 Db 1321 GTCTGCAACCCCGTGAAGGCGAAGGCAATGCCACCACTCCGTGAAGGTCCAGAGGCC 1380  
 QY 1466 CAGAGAGTCTTGGGGGACCCATGGGTGCGCATGGGAGCCGTGAGAGATAGAGAGGCT 1525  
 Db 1381 CAGAGAGTCTTGGGGGACCCATGGGTGCGCATGGGAGCCGTGAGAGATAGAGAGGCT 1440  
 QY 1526 TCCCTTCAAGAGGACCCCGGGGTTCCGAGGGAACCCCTCTGAGAGAGGAAAGTCCAGC 1585  
 Db 1441 TCCCTTCAAGAGGACCCCGGGGTTCCGAGGGAACCCCTCTGAGAGAGGAAAGTCCAGC 1500  
 QY 1586 ACCGAGCTGGAAGCTTGAGAGTGTCCCAATCCGGGTTGTGAAACAGAACTTCCCTGTGA 1645  
 Db 1501 ACCGAGCTGGAAGCTTGAGAGTGTCCCAATCCGGGTTGTGAAACAGAACTTCCCTGTGA 1560  
 QY 1646 GAAAGAGTGTGAAAGCTGTAAAGCCCAAGCTTAAAGGCTTCTCCCGCCACCTTTTC 1705  
 Db 1561 GAAAGAGTGTGAAAGCTGTAAAGCCCAAGCTTAAAGGCTTCTCCCGCCACCTTTTC 1620  
 QY 1706 CCCCCCGGGGTGTGATCGTGGCAGAGTGTGACCTGTGGCAATGGGGGAGGAAAGAGTG 1765  
 Db 1621 CCCCCCGGGGTGTGATCGTGGCAGAGTGTGACCTGTGGCAATGGGGGAGGAAAGAGTG 1680  
 QY 1766 GCTTCAATGATCGTGTGTGATCCCGGTGAATGTTCAAGAGTGTGCTGTGTGAGACAGAG 1825  
 Db 1681 GCTTCAATGATCGTGTGTGATCCCGGTGAATGTTCAAGAGTGTGCTGTGTGAGACAGAG 1740  
 QY 1826 GTGGCTTCCGTGTGTGATGGCCGGGGCAATGACCCGAGTGTGCTTTGTGTGAGAAACAGAGTG 1885  
 Db 1741 GTGGCTTCCGTGTGTGATGGCCGGGGCAATGACCCGAGTGTGCTTTGTGTGAGAAACAGAGTG 1800  
 QY 1886 GGCCTGTGGGGGGCCCCCTGAGACCTTTGATGAAACAGATGGGAGGAAAGAGAGAGCGTG 1945  
 Db 1801 GGCCTGTGGGGGGCCCCCTGAGACCTTTGATGAAACAGATGGGAGGAAAGAGAGAGCGTG 1860  
 QY 1946 GAGGACTGTGAAATAATGATAAGGCGAGACCTGTGAGAGGCGAGAGATCGGCTTACT 2005  
 Db 1861 GAGGACTGTGAAATAATGATAAGGCGAGACCTGTGAGAGGCGAGAGATCGGCTTACT 1920  
 QY 2006 AGATGCAAGAGACCCCGCAGA 2025  
 Db 1921 AGATGCAAGAGACCCCGCAGA 1940

RESULT 9  
 CR608666 2083 bp mRNA linear HTC 21-JUL-2004  
 LOCUS CR608666  
 DEFINITION Full-length cDNA clone CS0D6004YL23 of B cells (Ramos cell line) of  
 Homo sapiens (human).  
 ACCESSION CR608666

VERSION CR608666.1 GI:50489473  
 KEYWORDS HTC; CNSLT\_cDNA.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 REFERENCE 1 (bases 1 to 2083)  
 AUTHORS Li, W.B., Gruber, C., Jessee, J., and Polyes, D.  
 TITLE Full-length cDNA libraries and normalization  
 JOURNAL Unpublished  
 REMARK Contact : Feng Liang Email : fliang@lifetech.com URL :  
 http://fulllength.invitrogen.com/InvitrogenCorporation1600  
 Faraday Avenue  
 2 (bases 1 to 2083)  
 REFERENCE Genoscope.  
 AUTHORS Direct Submision  
 TITLE Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage :  
 BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr  
 JOURNAL - Web : www.genoscope.cns.fr)  
 COMMENT 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime  
 end enriched, double-strand cDNA was digested with Not I and cloned  
 into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library  
 was normalized. Library was constructed by Life Technologies, a  
 division of Invitrogen.  
 FEATURES Location/Qualifiers  
 source 1..2083  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="CS0D6004YL23"  
 /rissue\_type="B cells (Ramos cell line)"  
 /plasmid="pCMVSPORT\_6"  
 ORIGIN  
 Query Match 95.8%; Score 1940; DB 3; Length 2083;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 1940; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 86 GCTACAGTCTTACACCGCCAGCCCACTCAAGATATGACAGACACCCAGGCAATATG 145  
 Db 1 GCTACAGTCTTACACCGCCAGCCCACTCAAGATATGACAGACACCCAGGCAATATG 60  
 QY 146 GCGAACAAGCTATGAAACCTATGACACGCCCACTGATGCTATACCCAGGCTCAGA 205  
 Db 61 GCGAACAAGCTATGAAACCTATGACACGCCCACTGATGCTATACCCAGGCTCAGA 120  
 QY 206 CCACTGCAACTATGGGCGAGACCGCCTTATGCAACTCTTATGAGACAGCCCTCCACTGATT 265  
 Db 121 CCACTGCAACTATGGGCGAGACCGCCTTATGCAACTCTTATGAGACAGCCCTCCACTGATT 180  
 QY 266 ATACTACTCAACTGCCCCCAGGCAATGACAGCCAGCTGTCCAGGGGTATGGCACTGTG 325  
 Db 181 ATACTACTCAACTGCCCCCAGGCAATGACAGCCAGCTGTCCAGGGGTATGGCACTGTG 240  
 QY 326 CTTATGATACCACTGCTGTATACAGTCAACCAACCAAGGCTCTTATGAGAGTCACTG 385  
 Db 241 CTTATGATACCACTGCTGTATACAGTCAACCAACCAAGGCTCTTATGAGAGTCACTG 300  
 QY 386 CATATGGCACTAGCCTGTATCCAGCCTTATGGGAGAGGCGAGGCACTGACCTTA 445  
 Db 301 CATATGGCACTAGCCTGTATCCAGCCTTATGGGAGAGGCGAGGCACTGACCTTA 360  
 QY 446 CAAAGCCGAGATGAAACAAAGCCCACTGAGACTAGTCAACCTCAATATAGACAGGGG 505  
 Db 361 CAAAGCCGAGATGAAACAAAGCCCACTGAGACTAGTCAACCTCAATATAGACAGGGG 420  
 QY 506 GTTACCAACAGCCGCTTATGAGATATGACAGATGTAATCCAGTATCCAGTATCTG 565  
 Db 421 GTTACCAACAGCCGCTTATGAGATATGACAGATGTAATCCAGTATCCAGTATCTG 480  
 QY 566 GGAAGCTACCCCATGAGCAGTCACTGACCTCCATCTTACCTTACCTTACCTTATCT 625  
 Db 481 GGAAGCTACCCCATGAGCAGTCACTGACCTCCATCTTACCTTACCTTACCTTATCT 540

OY	626	CTACACAGCCGACATAGTATATATGATGAGACAGTACTCTGACAGCAAAACACCTATGCGCAAC	685
Db	541	CTTACACAGCCGACATAGTATATATGATGAGACAGTACTCTGACAGCAAAACACCTATGCGCAAC	600
OY	686	CGACGACGCTATGACAGCAGAGTAGCTATGCGTCAACAAAGCAGTATGCGCAGCAGCTC	745
Db	601	CGACGACGCTATGACAGCAGAGTAGCTATGCGTCAACAAAGCAGTATGCGCAGCAGCTC	660
OY	746	CCACTAGTATACCAACCCCAATCGGATGCTACACGCAAGCTCCAAAGTCAATATATGCGCAAC	805
Db	661	CCACTAGTATACCAACCCCAATCGGATGCTACACGCAAGCTCCAAAGTCAATATATGCGCAAC	720
OY	806	AGAGCAGCAGCTACCGGCGACGAGATTCAATCCGACAGAACCAACCCAGTAGAGTGGGTG	865
Db	721	AGAGCAGCAGCTACCGGCGACGAGATTCAATCCGACAGAACCAACCCAGTAGAGTGGGTG	780
OY	866	TTTATGCGCAGAGAGTCTGAGAGATTTTTCGGAACGAGAGAGAACCGGAGCATGAGTGGCC	925
Db	781	TTTATGCGCAGAGAGTCTGAGAGATTTTTCGGAACGAGAGAGAACCGGAGCATGAGTGGCC	840
OY	926	CTGATPACCGGGGCGAGGGGAGAGGGGGATTGTGATCGTAGAGCATGAGCAGAGSTGGGC	985
Db	841	CTGATPACCGGGGCGAGGGGAGAGGGGGATTGTGATCGTAGAGCATGAGCAGAGSTGGGC	900
OY	986	GGGGAGAGAGACGCGGTGAAATGGGACGCGCTGAGAGCAGAGGTGGCTTCATTAAGCCTG	1045
Db	901	GGGGAGAGAGACGCGGTGAAATGGGACGCGCTGAGAGCAGAGGTGGCTTCATTAAGCCTG	960
OY	1046	GTGGACCCATGGATGGAAGAACGACATCTTGTAACGTAGGCCCATCGTATGATCAATATAG	1105
Db	961	GTGGACCCATGGATGGAAGAACGACATCTTGTAACGTAGGCCCATCGTATGATCAATATAG	1020
OY	1106	ACTCTGACAAACAGTGCATTTATATGACAGAGATTAAATGACAGTGTGACTCTAGATGATC	1165
Db	1021	ACTCTGACAAACAGTGCATTTATATGACAGAGATTAAATGACAGTGTGACTCTAGATGATC	1080
OY	1166	TGGCAGACTTCTTTAAGCAGTGTGGGGTTGTTAAGATGAACAAGAACTGGGCAACCCA	1225
Db	1081	TGGCAGACTTCTTTAAGCAGTGTGGGGTTGTTAAGATGAACAAGAACTGGGCAACCCA	1140
OY	1226	TGATCCACATCTACCTGGACAAAGAAACAGGAAAGCCAAAGGCGATGCCACAGTGTCT	1285
Db	1141	TGATCCACATCTACCTGGACAAAGAAACAGGAAAGCCAAAGGCGATGCCACAGTGTCT	1200
OY	1286	ATGAAGACCCACCACTGCTCAAGGCTGCGCTGGAAATGCGTTGATGGGAAAGATTTCAG	1345
Db	1201	ATGAAGACCCACCACTGCTCAAGGCTGCGCTGGAAATGCGTTGATGGGAAAGATTTCAG	1260
OY	1346	GGAGCAAACTTAAATGCTCCCTGCTCGGAAAGAAAGCTCCAAATGAACGATATGGGGGTG	1405
Db	1261	GGAGCAAACTTAAATGCTCCCTGCTCGGAAAGAAAGCTCCAAATGAACGATATGGGGGTG	1320
OY	1406	GTCTGCCACCCCGTGAAGGCGAAGGCATGCCACCACTCCGTGGAGSTCCAGAGGCC	1465
Db	1321	GTCTGCCACCCCGTGAAGGCGAAGGCATGCCACCACTCCGTGGAGSTCCAGAGGCC	1380
OY	1466	CAGAGAGTCTCTGGGGGAGCCCATAGGTGCGATGGAGGCGCGTGGAGAGATATGAGAGGCT	1525
Db	1381	CAGAGAGTCTCTGGGGGAGCCCATAGGTGCGATGGAGGCGCGTGGAGAGATATGAGAGGCT	1440
OY	1526	TCCCTCTCAAGAGAGCCCGGGGGTTTCCGAGGGGAAACCCCTCTGGAGAGAGAAACGTCAGC	1585
Db	1441	TCCCTCTCAAGAGAGCCCGGGGGTTTCCGAGGGGAAACCCCTCTGGAGAGAGAAACGTCAGC	1500
OY	1586	ACCGAGCTGGAGACTGGCAGTGTCCCAATCCGGGTTTGAGAAACCAAGAACTTTCGCTGGA	1645
Db	1501	ACCGAGCTGGAGACTGGCAGTGTCCCAATCCGGGTTTGAGAAACCAAGAACTTTCGCTGGA	1560
OY	1646	GAAACGAGTGAACCAAGTGTAGAGGCCCAAAACCTGAAAGGCTTCTCCCGGCAACCTTTC	1705
Db	1561	GAAACGAGTGAACCAAGTGTAGAGGCCCAAAACCTGAAAGGCTTCTCCCGGCAACCTTTC	1620

QY	1706	CGCCCCGGGGTGGATCGTGGCAGAGTGGCCCTGGTGGCATGCGGGGAGGAAGAGTGTG	1765
Db	1621	CGCCCCGGGGTGGATCGTGGCAGAGTGGCCCTGGTGGCATGCGGGGAGGAAGAGTGTG	1680
QY	1766	GCCCTCATGAGATCGTGTGGTGTCCGGTGGAAATGTTACAGAGGTGGCCGTGTGTGAGACAGAG	1825
Db	1681	GCCTCATGAGATCGTGTGGTGTCCGGTGGAAATGTTACAGAGGTGGCCGTGTGTGAGACAGAG	1740
QY	1826	GTGGCTTCCGTGTGTGGCCGGGGCATGGACCGAGGTGGCTTTGGTGGAGGAAGACGAGGTG	1885
Db	1741	GTGGCTTCCGTGTGTGGCCGGGGCATGGACCGAGGTGGCTTTGGTGGAGGAAGACGAGGTG	1800
QY	1886	GCCCTGGGGGGCCCCCCTGGACCTTTGATGGAAACGATGGGAGGAAGGAAGGAGAGCGTG	1945
Db	1801	GCCCTGGGGGGCCCCCCTGGACCTTTGATGGAAACGATGGGAGGAAGGAAGGAGAGCGTG	1860
QY	1946	GAGGACCTGGAAAAATGATTAAGGCGAGCACCGTCAGAGCGCAGAGATCGGCCCTTACT	2005
Db	1861	GAGGACCTGGAAAAATGATTAAGGCGAGCACCGTCAGAGCGCAGAGATCGGCCCTTACT	1920
QY	2006	AGATGCAAGACCCCGCAGA 2025	
Db	1921	AGATGCAAGACCCCGCAGA 1940	

RESULT 10			
CR604639			
LOCUS	2236 bp	mrna	linear
DEFINITION	full-length cDNA clone C6SD013YN06 of T cells (jurkat cell line)		
	Cot 10-normalized of Homo sapiens (human) .		
CR604639			HTC 21-JUL-2004

ACCESSION	CR604639
VERSION	CR604639.1
KEYWORDS	HTC; CNSLT_cDNA.
SOURCE	Homo sapiens
ORGANISM	Homo sapiens (human)

REFERENCE  
Eukaryota; Metazoa: Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1. (bases 1 to 2236)

AUTHORS  
 TITLE  
 JOURNAL  
 REMARK  
 Unpublished  
 Contact: Feng Liang Email: [fliang@life.techn.com](mailto:fliang@life.techn.com) URL: <http://fulllength.invitrogen.com/> Invitrogen Corporation 1600

Faraday Avenue  
2 (bases 1 to 2236)

**JOURNAL TITLE** Genoscope - Centre National de Séquençage :  
**AUTHORS** Direct Submision  
**Submitted (20-JUL-2004)** Submitted (20-JUL-2004)

COMMENT

- Web : [www.genoscoper.cns.fr](http://www.genoscoper.cns.fr)  
BP 151 91006 EVRY cedex - FRANCE (E-mail : [bedel@genoscoper.cns.fr](mailto:bedel@genoscoper.cns.fr))

1st strand cDNA was primed with a NotI-oligo(dT) primer. Five primers  
were used. Family 1 showed cDNA was discarded with Not I and cloned

into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies, a division of Invitrogen.

FEATURES	Location/Qualifiers
source	1. .2236

```

/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/db_xref="ecoli:330000"

```

```

/cellname="C57BL/6J"
/tissue_type="T cells (Jurkat cell line) Coc
10-normalized"
/plasmid="pCMVSPORT_6"

```

ORIGIN	
Query Match	92.1%; Score 1866; DB 3; Length 2236;
Best Local Similarity	99.8%; Pred. No. 0;
Matches 1679; Conservative	0; Mismatches 0; Indels 3; Gaps 1

19 GAGGAGAAAGAGAAAATGGCGTCCACGATTAAGCTTACGCAAGCTCAGG 78  
 30 |||||  
 40 |||||  
 50 |||||  
 60 |||||  
 70 |||||  
 80 |||||  
 90 |||||  
 100 |||||  
 110 |||||  
 120 |||||  
 130 |||||  
 140 |||||  
 150 |||||  
 160 |||||  
 170 |||||  
 180 |||||  
 190 |||||  
 200 |||||  
 210 |||||  
 220 |||||  
 230 |||||  
 240 |||||  
 250 |||||  
 260 |||||  
 270 |||||  
 280 |||||  
 290 |||||  
 300 |||||  
 310 |||||  
 320 |||||  
 330 |||||  
 340 |||||  
 350 |||||  
 360 |||||  
 370 |||||  
 380 |||||  
 390 |||||  
 400 |||||  
 410 |||||  
 420 |||||  
 430 |||||  
 440 |||||  
 450 |||||  
 460 |||||  
 470 |||||  
 480 |||||  
 490 |||||  
 500 |||||  
 510 |||||  
 520 |||||  
 530 |||||  
 540 |||||  
 550 |||||  
 560 |||||  
 570 |||||  
 580 |||||  
 590 |||||  
 600 |||||  
 610 |||||  
 620 |||||  
 630 |||||  
 640 |||||  
 650 |||||  
 660 |||||  
 670 |||||  
 680 |||||  
 690 |||||  
 700 |||||  
 710 |||||  
 720 |||||  
 730 |||||  
 740 |||||  
 750 |||||  
 760 |||||  
 770 |||||  
 780 |||||  
 790 |||||  
 800 |||||  
 810 |||||  
 820 |||||  
 830 |||||  
 840 |||||  
 850 |||||  
 860 |||||  
 870 |||||  
 880 |||||  
 890 |||||  
 900 |||||  
 910 |||||  
 920 |||||  
 930 |||||  
 940 |||||  
 950 |||||  
 960 |||||  
 970 |||||  
 980 |||||  
 990 |||||  
 1000 |||||  
 1010 |||||  
 1020 |||||  
 1030 |||||  
 1040 |||||  
 1050 |||||  
 1060 |||||  
 1070 |||||  
 1080 |||||  
 1090 |||||  
 1100 |||||  
 1110 |||||  
 1120 |||||  
 1130 |||||  
 1140 |||||  
 1150 |||||  
 1160 |||||  
 1170 |||||  
 1180 |||||  
 1190 |||||  
 1200 |||||  
 1210 |||||  
 1220 |||||  
 1230 |||||  
 1240 |||||  
 1250 |||||  
 1260 |||||  
 1270 |||||  
 1280 |||||  
 1290 |||||  
 1300 |||||  
 1310 |||||  
 1320 |||||  
 1330 |||||  
 1340 |||||  
 1350 |||||  
 1360 |||||  
 1370 |||||  
 1380 |||||  
 1390 |||||  
 1400 |||||  
 1410 |||||  
 1420 |||||  
 1430 |||||  
 1440 |||||  
 1450 |||||  
 1460 |||||  
 1470 |||||  
 1480 |||||  
 1490 |||||  
 1500 |||||  
 1510 |||||  
 1520 |||||  
 1530 |||||  
 1540 |||||  
 1550 |||||  
 1560 |||||  
 1570 |||||  
 1580 |||||  
 1590 |||||  
 1600 |||||  
 1610 |||||  
 1620 |||||  
 1630 |||||  
 1640 |||||  
 1650 |||||  
 1660 |||||  
 1670 |||||  
 1680 |||||  
 1690 |||||  
 1700 |||||  
 1710 |||||  
 1720 |||||  
 1730 |||||  
 1740 |||||  
 1750 |||||  
 1760 |||||  
 1770 |||||  
 1780 |||||  
 1790 |||||  
 1800 |||||  
 1810 |||||  
 1820 |||||  
 1830 |||||  
 1840 |||||  
 1850 |||||  
 1860 |||||  
 1870 |||||  
 1880 |||||  
 1890 |||||  
 1900 |||||  
 1910 |||||  
 1920 |||||  
 1930 |||||  
 1940 |||||  
 1950 |||||  
 1960 |||||  
 1970 |||||  
 1980 |||||  
 1990 |||||  
 2000 |||||  
 2010 |||||  
 2020 |||||  
 2030 |||||  
 2040 |||||  
 2050 |||||  
 2060 |||||  
 2070 |||||  
 2080 |||||  
 2090 |||||  
 2100 |||||  
 2110 |||||  
 2120 |||||  
 2130 |||||  
 2140 |||||  
 2150 |||||  
 2160 |||||  
 2170 |||||  
 2180 |||||  
 2190 |||||  
 2200 |||||  
 2210 |||||  
 2220 |||||  
 2230 |||||  
 2240 |||||  
 2250 |||||  
 2260 |||||  
 2270 |||||  
 2280 |||||  
 2290 |||||  
 2300 |||||  
 2310 |||||  
 2320 |||||  
 2330 |||||  
 2340 |||||  
 2350 |||||  
 2360 |||||  
 2370 |||||  
 2380 |||||  
 2390 |||||  
 2400 |||||  
 2410 |||||  
 2420 |||||  
 2430 |||||  
 2440 |||||  
 2450 |||||  
 2460 |||||  
 2470 |||||  
 2480 |||||  
 2490 |||||  
 2500 |||||  
 2510 |||||  
 2520 |||||  
 2530 |||||  
 2540 |||||  
 2550 |||||  
 2560 |||||  
 2570 |||||  
 2580 |||||  
 2590 |||||  
 2600 |||||  
 2610 |||||  
 2620 |||||  
 2630 |||||  
 2640 |||||  
 2650 |||||  
 2660 |||||  
 2670 |||||  
 2680 |||||  
 2690 |||||  
 2700 |||||  
 2710 |||||  
 2720 |||||  
 2730 |||||  
 2740 |||||  
 2750 |||||  
 2760 |||||  
 2770 |||||  
 2780 |||||  
 2790 |||||  
 2800 |||||  
 2810 |||||  
 2820 |||||  
 2830 |||||  
 2840 |||||  
 2850 |||||  
 2860 |||||  
 2870 |||||  
 2880 |||||  
 2890 |||||  
 2900 |||||  
 2910 |||||  
 2920 |||||  
 2930 |||||  
 2940 |||||  
 2950 |||||  
 2960 |||||  
 2970 |||||  
 2980 |||||  
 2990 |||||  
 3000 |||||  
 3010 |||||  
 3020 |||||  
 3030 |||||  
 3040 |||||  
 3050 |||||  
 3060 |||||  
 3070 |||||  
 3080 |||||  
 3090 |||||  
 3100 |||||  
 3110 |||||  
 3120 |||||  
 3130 |||||  
 3140 |||||  
 3150 |||||  
 3160 |||||  
 3170 |||||  
 3180 |||||  
 3190 |||||  
 3200 |||||  
 3210 |||||  
 3220 |||||

QY 79 CAGCAGGCGCTACAGTGTCTTACACCGCCAGCCCACTGAAGATATGACAGACCAACCAG 138  
 DB 61 CAGCAGGCGCTACAGTGTCTTACACCGCCAGCCCACTGAAGATATGACAGACCAACCAG 120  
 QY 139 GCATATGGGCAACAAAGCTATGGAACCTATGAGCAGCCCACTGATGTCACTATATACCAG 198  
 DB 121 GCATATGGGCAACAAAGCTATGGAACCTATGAGCAGCCCACTGATGTCACTATATACCAG 180  
 QY 199 GCTCAGACCACTGCAACCTATGAGGAGACGCGCTATGCAACTTCTTATGAGCAGCTCCC 258  
 DB 181 GCTCAGACCACTGCAACCTATGAGGAGACGCGCTATGCAACTTCTTATGAGCAGCTCCC 240  
 QY 259 ACTGCTTATATCTCTCCAACTGCCCCCAGGCAATACAGCCAGCTGTCCAGGGGTATGCG 318  
 DB 241 ACTGCTTATATCTCTCCAACTGCCCCCAGGCAATACAGCCAGCTGTCCAGGGGTATGCG 300  
 QY 319 ACTGCTTATATCTCTCCAACTGCCCCCAGGCAATACAGCCAGCTGTCTTATGAGCAGCT 378  
 DB 301 ACTGCTTATATCTCTCCAACTGCCCCCAGGCAATACAGCCAGCTGTCTTATGAGCAGCT 360  
 QY 379 CAGTCTGCAATATGCACTCAGCTGCTTATTCAGCTTATGAGGAGCAGCCAGCAGCACT 438  
 DB 361 CAGTCTGCAATATGCACTCAGCTGCTTATTCAGCTTATGAGGAGCAGCCAGCAGCACT 420  
 QY 439 GCACCTACAGACCCGAGAGTGAACAAAGCCCACTGAGACTGATCACTCAATCTAGC 498  
 DB 421 GCACCTACAGACCCGAGAGTGAACAAAGCCCACTGAGACTGATCACTCAATCTAGC 480  
 QY 499 ACAGGGGGTTTCAACACGAGCCAGCTAGATATGAGACAGAGTAACTACAGTTATCCCGAG 558  
 DB 481 ACAGGGGGTTTCAACACGAGCCAGCTAGATATGAGACAGAGTAACTACAGTTATCCCGAG 540  
 QY 559 GTACCTGGAGGCTACCCCACTGAGCCAGCTCAGCTCCTCACTACCTCTTACAGC 618  
 DB 541 GTACCTGGAGGCTACCCCACTGAGCCAGCTCAGCTCCTCACTACCTCTTACAGC 600  
 QY 619 TATTTCTCTACAGACCCGAGCTATGATATGAGCAGCTTACTCTCAGCAGAACCTAT 678  
 DB 601 TATTTCTCTACAGACCCGAGCTATGATATGAGCAGCTTACTCTCAGCAGAACCTAT 660  
 QY 679 GGGCAACCGAGCAGCTATGAGCAGCAGAGTATGATGATCAACAAAGCAGCTATGAGGAG 738  
 DB 661 GGGCAACCGAGCAGCTATGAGCAGCAGAGTATGATGATCAACAAAGCAGCTATGAGGAG 720  
 QY 739 CAGCTTCCCACTAGTTACCCCACTGATTCCTACAGCCAGCTCCAGTCAATAT 798  
 DB 721 CAGCTTCCCACTAGTTACCCCACTGATTCCTACAGCCAGCTCCAGTCAATAT 780  
 QY 799 AGCCAAACAGAGCAGCAGCTACGGGAGCAGAGATTCAATCCGACAGACCAACCCAGTAGC 858  
 DB 781 AGCCAAACAGAGCAGCAGCTACGGGAGCAGAGATTCAATCCGACAGACCAACCCAGTAGC 840  
 QY 859 ATGGGTGTTTATGGGAGAGAGTCTGAGAGATTTTCGGGACAGAGAGAACCCGAGCAGT 918  
 DB 841 ATGGGTGTTTATGGGAGAGAGTCTGAGAGATTTTCGGGACAGAGAGAACCCGAGCAGT 900  
 QY 919 AGTGGCCCTGATTAACCGGGGAGGAGGAGATTGATGCTGAGAGCAGTATGAGCAG 978  
 DB 901 AGTGGCCCTGATTAACCGGGGAGGAGGAGATTGATGCTGAGAGCAGTATGAGCAG 960  
 QY 979 GGTGGGCGGGAGAGAGAGCGCGGTGGAATGGGAGCGGCTGAGAGAGAGGTGCTCAAT 1038  
 DB 961 GGTGGGCGGGAGAGAGAGCGCGGTGGAATGGGAGCGGCTGAGAGAGAGGTGCTCAAT 1017  
 QY 1039 AAGCCTGTGAGACCCATGATGAAGGACAGATCTTATGATAGGCCCACTGTATGATCA 1098  
 DB 1018 AAGCCTGTGAGACCCATGATGAAGGACAGATCTTATGATAGGCCCACTGTATGATCA 1077  
 QY 1099 GATGAAGACTGTGACACAGTGAATTTATGTAACAAGATTAATGACAGTGTACTCTTA 1158  
 DB 1078 GATGAAGACTGTGACACAGTGAATTTATGTAACAAGATTAATGACAGTGTACTCTTA 1137

QY 1159 GATGATCTGGCAGACCTTTTAAAGCAGTGTGGGTTTGAATGAACAAGAACCTGG 1218  
 DB 1138 GATGATCTGGCAGACCTTTTAAAGCAGTGTGGGTTTGAATGAACAAGAACCTGG 1197  
 QY 1219 CAACCCATGATCAACATCTACCTGAGCAAGAAACAGAAAGCCCAAGGCGATGCCA 1278  
 DB 1198 CAACCCATGATCAACATCTACCTGAGCAAGAAACAGAAAGCCCAAGGCGATGCCA 1257  
 QY 1279 GTGTCTTATGAAGACCAACCACTGCCCAAGGCTGCGGTGAAATGTTTGAATGGAAAGAT 1338  
 DB 1258 GTGTCTTATGAAGACCAACCACTGCCCAAGGCTGCGGTGAAATGTTTGAATGGAAAGAT 1317  
 QY 1339 TTTCAGAGGAGCAAACTTAAAGTCTCCCTTGTCTGGAAGAACCTCCATGAACAGTATG 1398  
 DB 1318 TTTCAGAGGAGCAAACTTAAAGTCTCCCTTGTCTGGAAGAACCTCCATGAACAGTATG 1377  
 QY 1399 CGGGGTGTGTGTCACCCGCTGAGGGGAGAGGAGATGCAACCACTCCGTTGAGAGTCCA 1458  
 DB 1378 CGGGGTGTGTGTCACCCGCTGAGGGGAGAGGAGATGTCACCACTCCGTTGAGAGTCCA 1437  
 QY 1459 GAGGCCCAAGAGGTCCTGAGGAGACCAATGGGTGCAATGAGAGGCGTGAAGAGATAGA 1518  
 DB 1438 GAGGCCCAAGAGGTCCTGAGGAGACCAATGGGTGCAATGAGAGGCGTGAAGAGATAGA 1497  
 QY 1519 GAGGCTTCCCTTCAAGAGGACCCCGGGGTTCCCGAGGGAACCCCTTGAAGAGGAAC 1578  
 DB 1498 GAGGCTTCCCTTCAAGAGGACCCCGGGGTTCCCGAGGGAACCCCTTGAAGAGGAAC 1557  
 QY 1579 GTCACGACCCGAGCTGAGAGCTGAGCAGTGTCCCAATCCGGGTTGTGAACCAAGATTC 1638  
 DB 1558 GTCACGACCCGAGCTGAGAGCTGAGCAGTGTCCCAATCCGGGTTGTGAACCAAGATTC 1617  
 QY 1639 GCCTGAGAAACAGAGTGCAACCAAGTGAAGGCCCAAGCCTGAAGGCTTCTCCGCCA 1698  
 DB 1618 GCCTGAGAAACAGAGTGCAACCAAGTGAAGGCCCAAGCCTGAAGGCTTCTCCGCCA 1677  
 QY 1699 CCTTTTCCGCCCCCGGGTGTGATCGTGGCAGAGGTGGCCCTGTGTCATGCCGGAGGA 1758  
 DB 1678 CCTTTTCCGCCCCCGGGTGTGATCGTGGCAGAGGTGGCCCTGTGTCATGCCGGAGGA 1737  
 QY 1759 AGAGGTGGCCTCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1818  
 DB 1738 AGAGGTGGCCTCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1797  
 QY 1819 GACAGAGGTGCTTCCGTGTGTCGCGGGGATGAGCCGAGGTGCTTGTGTGAGAGAGA 1878  
 DB 1798 GACAGAGGTGCTTCCGTGTGTCGCGGGGATGAGCCGAGGTGCTTGTGTGAGAGAGA 1857  
 QY 1879 CGAGGTGGCCTTGGGGGGCCCC 1900  
 DB 1858 CGAGGTGGCCTTGGGGGGCCCC 1879

RESULT 11  
 AK034755  
 LOCUS 2373 bp mRNA linear HTC 03-APR-2004  
 DEFINITION Mus musculus 12 days embryo embryonic body between diaphragm region  
 and neck cDNA, RIKEN full-length enriched library, clone:943003M11  
 product:Ewing sarcoma homolog, full insert sequence.  
 AK034755.1 GI:26084174  
 VERSION AK034755.1  
 KEYWORDS HTC; GAP trapper.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE  
 1 Carninci, P. and Hayashizaki, Y.  
 High-efficiency full-length cDNA cloning  
 Meth. Enzymol. 303, 19-44 (1999)  
 JOURNAL MEDLINE 99279253  
 PUBMED 10349636  
 REFERENCE 2



AUTHORS	Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
TITLE	Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL	Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE	20493974
PUBMED	11042159
REFERENCE	
AUTHORS	3 Shibata, K., Itoh, M., Aizawa, K., Nagao, S., Sasaki, N., Carninci, P., Kono, H., Akiyama, U., Nishi, K., Kitenaka, T., Tashiro, H., Itoh, M., Smit, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Harada, A., Yamamoto, R., Matsunoto, H., Sakaguchi, S., Ikegami, T., Kasai, K., Fujiwara, K., Inoue, K., Togawa, Y., Iwata, M., Ohara, E., Watanabe, Y., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kiz, A. and Hayashizaki, Y.
TITLE	RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer
JOURNAL	Genome Res. 10 (11), 1757-1771 (2000)
MEDLINE	20530913
PUBMED	11076861
REFERENCE	
AUTHORS	4 The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.
TITLE	Functional annotation of a full-length mouse cDNA collection
JOURNAL	Nature 409, 685-690 (2001)
MEDLINE	
PUBMED	
REFERENCE	
AUTHORS	5 The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.
TITLE	Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
JOURNAL	Nature 420, 563-573 (2002)
MEDLINE	6 (bases 1 to 2373)
PUBMED	
REFERENCE	
AUTHORS	Adachi, S., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fuchida, S., Furuta, M., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hizemoto, K., Hirooka, T., Hirozane, T., Hori, F., Imocani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Karch, H., Kawai, J., Kojima, Y., Kondo, S., Kono, H., Konda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishii, K., Nomura, K., Nunazaki, R., Ohno, M., Ohsato, N., Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shingawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.
TITLE	Direct Submission
JOURNAL	Submitted (16-JUL-2001) Yoshinide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama Kanagawa 230-0045, Japan (E-mail: genome-research@riken.jp, URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)
COMMENT	cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in Riken Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. Please visit our web site for further details. URL: http://genome.gsc.riken.jp/ URL: http://fantom.gsc.riken.jp/. Location/Qualifiers
FEATURES	1. 2373
SOURCE	1. Organism="Mus musculus" 2. mol_type="mRNA" 3. strain="C57BL/6J" 4. db_xref="FANTOM,DB:9430031M11" 5. db_xref="taxon:10090" 6. clone="9430031M11" 7. tissue_type="embryonic body between diaphragm region and neck" 8. clone_jib="RIKEN full-length enriched mouse cDNA library 9. dev_stage="12 days embryo" 10. misc_feature

[illegible]



QY 900 AGAGAGAACCGGAGCATGAGTGGCCCTGATTAACCGGGGCAAGGGGAAAGGGGGATTGA 959  
 DB 907 AGAGAGAACCGGAGCATGAGTGGCCCTGATTAACCGGGGCAAGGGGAAAGGGGGATTGA 966  
 QY 960 TCGTGAAGCATGAGCAGAGGTGGCGGGGAGAGAGACGCGGTGAATGGGACCGCTGG 1019  
 DB 967 TCGTGAAGCATGAGCAGAGGTGGCGGGGAGAGAGACGCGGTGAATGGGACCGCTGG 1023  
 QY 1020 AGAGCAGAGGTGGCTTCAATTAAGCCTGGTGAACCAATGATGAAGAGACCAATCTTGAATCT 1079  
 DB 1024 AGAGCAGAGGTGGCTTCAATTAAGCCTGGTGAACCAATGATGAAGAGACCAATCTTGAATCT 1083  
 QY 1080 AGAGCCCACTGTAGATCCAGATGAAGACTGTGAACAAGTGAATTTATGTAACAAGATTT 1139  
 DB 1084 AGAGCTTCTATATATCCCAATGAAGACTGTGAACAAGTGAATTTATGTAACAAGATTT 1143  
 QY 1140 AAATGACAGTGTGACTTATAGTATCTTGGCAGACTTCTTTAAGCAGTGTGGGTTGTTAA 1199  
 DB 1144 AAATGACAGTGTGACTTATAGTATCTTGGCAGACTTCTTTAAGCAGTGTGGGTTGTTCA 1203  
 QY 1200 GATGAACAAGAACTGGGCAACCCATGATTCACATCTTACCTGGAACAAGAAACAGAA 1259  
 DB 1204 GATGAACAAGAACTGGGCAACCCATGATTCACATCTTACCTGGAACAAGAAACAGAA 1263  
 QY 1260 GCCCAAGGGGATGCCAAGTGTCTATGAAGACCCCACTGCCCAAGGCTGCCGTGA 1319  
 DB 1264 GCCCAAGGGGATGCCAAGTGTCTATGAAGATTCACCACTGGAAGGCGAGGAGTCCACC 1323  
 QY 1320 ATGTTTGAATGAGAAATTTTCAAGGAGCAAACTTAAAGTCTCCCTGTCTGGAAGAA 1379  
 DB 1324 ATGTTTGAATGAGAAATTTTCAAGGAGCAAACTTAAAGTCTCTTCCGCAAAAGAA 1383  
 QY 1380 GCTCTCAATGAACATATGCGGGGTGTCTGCAACCCCGTGAAGGCAAGGCGATGCCACC 1439  
 DB 1384 GCTCTCAATGAACATATGCGGGGTGTCTGCAACCCCGTGAAGGCGAGGAGTCCACC 1443  
 QY 1440 ACCACTCGTGAAGTTCAGAGGCGCAAGAGTCTCGGGGAGCCCAATGGTGTGCATAGGG 1499  
 DB 1444 ACCACTCGTGAAGTTCAGAGGCGCAAGAGTCTCGGGGAGCCCAATGGTGTGCATAGGG 1503  
 QY 1500 AGGCGGTGAAGATGAGAGAGCTTCCCTCCAAAGAGACCCCGGGGTTCCCGAGGAGAA 1559  
 DB 1504 AGGCGGTGAAGATGAGAGAGCTTCCCTCCAAAGAGAGCCCGGGGTTCCCGAGGAGAA 1563  
 QY 1560 CCCCTCTGAGAGAGAAACGTCAGACCCAGAGTGAAGATGTGCCAATCCGAGG 1619  
 DB 1564 CCCCTCTGAGAGAGAAACGTCAGACCCAGAGTGAAGATGTGCCAATCCGAGG 1623  
 QY 1620 TTGTGGAACACAGAACTTCCCTGAGAGACAGAGTGAACCAAGTGAAGGCCCCCAAGCC 1679  
 DB 1624 TTGTGGAACACAGAACTTCCCTGAGAGACAGAGTGAACCAAGTGAAGGCCCCCAAGCC 1683  
 QY 1680 TGAAGGCTTCTCCCGCCACCTTTCCGCCCGGGGTGTGATGTGGCAGAGGTGGCC 1739  
 DB 1684 TGAAGGCTTCTCCCGCCACCTTTCCGCCCGGGGTGTGATGTGGCAGAGGTGGCC 1743  
 QY 1740 TGGTGAATCGGGGAGAAAGAGTGGCTTATGATGTGTGTGTCCGCTGGAATGTT 1799  
 DB 1744 TGGTGAATCGGGGAGAAAGAGTGGCTTATGATGTGTGTGTCCGCTGGAATGTT 1803  
 QY 1800 CAAAGGTGGCTGTGTGAGACAGAGTGTGCTTCCGTGTGTGCGGGGCAATGACAGAG 1859  
 DB 1804 CAAAGGTGGCTGTGTGAGACAGAGTGTGCTTCCGTGTGTGCGGGGCAATGACAGAG 1863  
 QY 1860 TGGCTTTGTGAGAGAAAGAGTGGCTTGGGGGGGCCCTGTGACCTTTTGAATGAACA 1919  
 DB 1864 TGGCTTTGTGAGAGAAAGAGTGGCTTGGGGGGGCCCTGTGACCTTTTGAATGAACA 1923  
 QY 1920 GATGGAGAGAAAGAGAGAGAGTGTGAGAGCTTGAAGAAATGATTAAGCGAGCACCG 1979  
 DB 1924 GATGGAGAGAAAGAGAGAGAGTGTGAGAGCTTGAAGAAATGATTAAGCGAGCACCG 1983  
 QY 1980 TCAGAGGCGAGAGATGGCCCTACTATGA 2008

Db 1984 TCAGAGGCGAGAGATGGCCCTACTATGA 2012  
 RESULT 12  
 AK019460  
 LOCUS  
 DEFINITION  
 AK019460 2107 bp mRNA linear HTC 03-APR-2004  
 Mus musculus 18 days pregnant adult female placenta and extra  
 embryonic tissue cDNA, RIKEN full-length enriched library,  
 clone:3830417B11 product:Ewing sarcoma homolog, full insert  
 sequence.  
 AK019460  
 VERSION  
 AK019460.1 GI:12859676  
 KEYWORDS  
 HTC; CAP trapper.  
 SOURCE  
 Mus musculus (house mouse)  
 ORGANISM  
 Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 REFERENCES  
 1 Carninci, P. and Hayashizaki, Y.  
 High-efficiency full-length cDNA cloning  
 Meth. Enzymol. 303, 19-44 (1999)  
 99279253  
 MEDLINE  
 PUBMED  
 10349636  
 2 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,  
 Itoh, M., Komno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
 Normalization and subtraction of cap-trapper-selected cDNAs to  
 prepare full-length cDNA libraries for rapid discovery of new genes  
 Genome Res. 10 (10), 1617-1630 (2000)  
 20499374  
 MEDLINE  
 PUBMED  
 11042159  
 3 Shibata, K., Itoh, M., Aizawa, K., Nagao, S., Sasaki, N., Carninci, P.,  
 Komno, H., Akiyama, J., Nishi, K., Kishimoto, T., Tashiro, H., Itoh, M.,  
 Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,  
 Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,  
 Fujiwara, S., Inoue, K., Togawa, Y., Iwata, M., Ohara, E., Watanabe, M.,  
 Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsunaga, S., Kawai, J.,  
 Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.  
 RIKEN integrated sequence analysis (RISA) system-384-format  
 sequencing pipeline with 384 multicapillary sequencer  
 Genome Res. 10 (11), 1757-1771 (2000)  
 20530913  
 MEDLINE  
 PUBMED  
 11076861  
 4 The RIKEN Genome Exploration Research Group Phase II Team and the  
 PANTOM Consortium.  
 Functional annotation of a full-length mouse cDNA collection  
 Nature 409, 685-690 (2001)  
 5 The FANTOM Consortium and the RIKEN Genome Exploration Research  
 Group Phase I & II Team.  
 Analysis of the mouse transcriptome based on functional annotation  
 of 60,770 full-length cDNAs  
 Nature 420, 563-573 (2002)  
 6 (bases 1 to 2107)  
 REFERENCE  
 AUTHORS  
 ADACHI, J., AIZAWA, K., AKIHARA, S., AKIMURA, T., ARAI, A., AONO, H.,  
 ARAKAWA, T., BONO, H., CARNINCI, P., FUKUNISHI, Y.,  
 HIRAKAWA, T., HORI, F., IMOCANI, K., ISHII, Y., KOMNO, H., KOWADA, M.,  
 KASUKAWA, T., KATO, H., KAWAI, J., KOJIMA, Y., KOMNO, H., KOWADA, M.,  
 KODA, S., KURIHARA, C., MATSUYAMA, T., MIYAZAKI, A., NISHI, K.,  
 NOMURA, K., NUMAZAKI, R., OHNO, M., OKAZAKI, Y., OKIDO, T., OWA, C.,  
 SAITO, H., SAITO, R., SAKAI, C., SAKAI, S., SANO, H., SAKAKI, D.,  
 SHIBATA, K., SHIBATA, Y., SHINAGAWA, A., SHIRAKI, T., SOGABE, Y.,  
 SUZUKI, H., TAGAMI, M., TAGAWA, A., TAKAHASHI, F., TANAKA, T.,  
 TEJIMA, Y., TOYA, T., YAMAMURA, T., YASUNISHI, A., YOSHIDA, K.,  
 YOSHINO, M., MURAMATSU, M. and Hayashizaki, Y.  
 Direct SubMISSION  
 Submitted (18-AUG-2000) Yoshihide Hayashizaki, The Institute of  
 Physical and Chemical Research (RIKEN), Laboratory for Genome  
 Exploration Research Group, RIKEN Genomic Sciences Center (GSC),



QY 1341 TCAAGGAGCAAACTTAAGTCTCCCTTCTCGAAGAGAGCTTCAATGAACAGTATGCG 1400  
 DB 1318 TCAAGGAGCAAACTTAAGTCTCTTCTCGAAGAGAGCTTCAATGAACAGTATGCG 1377  
 QY 1401 GGGTGTCTCCCAACCCCGTAGAGGAGGAGGATGCGACCACTCCGTGAGAGGTCCAGG 1460  
 DB 1378 GGGAGGATATCCCACTCTGTAGAGGAGGAGGATGCGACCACTCTGTGAGAGGTCTG 1437  
 QY 1461 AGGCGCCAGAGAGTCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1520  
 DB 1438 TGGCCCGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1497  
 QY 1521 AGGCTTCCCTTCAAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1580  
 DB 1498 GGGCTTCTCTTCAAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1557  
 QY 1581 CCAGACCCGAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 1640  
 DB 1558 CCAGACCCGAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 1617  
 QY 1641 CTGAGAGAGAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 1700  
 DB 1618 TTGAGAGAGAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 1677  
 QY 1701 CTTTCCGCCCCCGGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 1760  
 DB 1678 CTTTCCGCCCCCGGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 1737  
 QY 1761 AGGTGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 1820  
 DB 1738 AGGTGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 1797  
 QY 1821 CAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 1880  
 DB 1798 CAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 1857  
 QY 1881 AGGTGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 1940  
 DB 1858 AGGTGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 1917  
 QY 1941 AGGTGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 2000  
 DB 1918 AGGTGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 1977  
 QY 2001 CTACTAGA 2008  
 DB 1978 CTACTAGA 1985  
 RESULT 13  
 AK049743  
 LOCUS AK049743 2269 bp mRNA 1linear HTC 03-Apr-2004  
 DEFINITION Mus musculus 12 days embryo spinal cord cDNA, RIKEN full-length  
 enriched library, clone: C530046A18 product: Bwling sarcoma homolog,  
 full insert sequence.  
 AK049743  
 VERSION AK049743.1 GI:26093614  
 KEYWORDS HTC, CAP trapper.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 REFERENCE  
 AUTHORS Carninci, P. and Hayashizaki, Y.  
 TITLE High-efficiency full-length cDNA cloning  
 JOURNAL Meth. Enzymol. 303, 19-44 (1999)  
 MEDLINE 99279253  
 PUBMED 10349636  
 REFERENCE  
 AUTHORS Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,  
 Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
 TITLE Normalization and subtraction of cap-trapper-selected cDNAs to

JOURNAL  
 MEDLINE 20499374  
 PUBMED 11042159  
 REFERENCE  
 AUTHORS  
 3  
 Shibata, K., Itoh, M., Aizawa, K., Nagaoaka, S., Sasaki, N., Carninci, P.,  
 Kono, H., Miyama, J., Nishii, K., Katsunai, T., Tashiro, H., Itoh, M.,  
 Sumi, N., Iehli, Y., Nakamura, S., Harada, M., Nishine, T., Harada, A.,  
 Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,  
 Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watanabe, M.,  
 Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuda, S., Kawai, J.,  
 Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.  
 RIKEN integrated sequence analysis (RISA) system-384-format  
 sequencing pipeline with 384 multicapillary sequencer  
 Genome Res. 10 (11), 1757-1771 (2000)  
 JOURNAL  
 MEDLINE 20530913  
 PUBMED 11076861  
 REFERENCE  
 AUTHORS  
 4  
 The RIKEN Genome Exploration Research Group Phase II Team and the  
 FANTOM Consortium.  
 Functional annotation of a full-length mouse cDNA collection  
 Nature 409, 685-690 (2001)  
 JOURNAL  
 MEDLINE 11076861  
 REFERENCE  
 AUTHORS  
 5  
 The FANTOM Consortium and the RIKEN Genome Exploration Research  
 Group Phase I & II Team.  
 Analysis of the mouse transcriptome based on functional annotation  
 of 60,770 full-length cDNAs  
 Nature 420, 563-573 (2002)  
 JOURNAL  
 MEDLINE 11076861  
 REFERENCE  
 AUTHORS  
 6  
 (bases 1 to 2269)  
 Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P.,  
 Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W.,  
 Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T.,  
 Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T.,  
 Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Kono, H., Kouda, M.,  
 Koya, S., Kuribara, C., Matsuyama, T., Miyazaki, A., Murata, M.,  
 Nakamura, Y., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N.,  
 Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N.,  
 Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T.,  
 Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S.,  
 Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A.,  
 Muramatsu, M. and Hayashizaki, Y.  
 Direct Submision  
 Submitted (16-Jul-2001) Yoshihide Hayashizaki, The Institute of  
 Physical and Chemical Research (RIKEN), Laboratory for Genome  
 Exploration and Research Group, RIKEN Genomic Sciences Center (GSC),  
 RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Teurumi-ku, Yokohama,  
 Kanagawa 230-0045, Japan (E-mail: genome-res@gs.riken.jp,  
 URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222,  
 Fax: 81-45-503-9216)  
 CDNA library was prepared and sequenced in Mouse Genome  
 Encyclopedia Project of Genome Exploration Research Group in Riken  
 Genomic Sciences Center and Genome Science Laboratory in RIKEN.  
 Division of Experimental Animal Research in Riken contributed to  
 prepare mouse tissues.  
 Please visit our web site for further details.  
 URL: http://genome.gsc.riken.jp/  
 URL: http://fantom.gsc.riken.jp/  
 Location/Qualifiers  
 1..2269  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="C57BL/6J"  
 /db\_xref="FANTOM DB:C530046A18"  
 /db\_xref="taxon:10090"  
 /clone="C530046A18"  
 /issue\_type="spinal cord"  
 /clone\_id="RIKEN full-length enriched mouse cDNA library"  
 /dev\_stage="12 days embryo"  
 48..1903  
 /note="Bwling sarcoma homolog (MGD)MGI:99960, GB|NM\_007968,  
 evidence: BIASTV, 99%, match=2172)  
 putative"  
 2251..2256  
 polyA\_signal



RESULT 14  
BX325713  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLES  
JOURNAL  
COMMENT

BX325713 1049 bp mRNA linear EST 08-APR-2004  
BX325713 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA  
clone CS0D1039YJ24 5-PRIME, mRNA sequence.  
BX325713  
BX325713.2 GI:46280663  
EST.  
Homo sapiens (human)  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 1049)  
L1.W.B., Gruber,C., Jessee,J. and Polyes,D.  
Full-length cDNA libraries and normalization  
unpublished (2001)  
On May 2, 2003 this sequence version replaced gi:30344479.  
Contact: Genoscope  
Genoscope - Centre National de Sequencage  
2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE  
Email: segre@genoscope.cns.fr, Web : www.genoscope.cns.fr  
1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime  
end enriched, double-strand cDNA was digested with Not I and cloned  
into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library  
was normalized. Library was constructed by Life Technologies, a  
division of Invitrogen. This sequence belongs to sequence cluster  
10589.f  
For more information about this cluster, see  
http://www.genoscope.cns.fr/cdna?c=CS0S012ZH02QP1&c=10589.f.  
Location/Qualifiers  
1..1049  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="CS0D1039YJ24"  
/issue\_type="PLACENTA COT 25-NORMALIZED"  
/clone\_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"  
/note="1st strand cDNA was primed with a NotI-oligo(dT)  
primer. Five prime end enriched, double-strand cDNA was  
digested with Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN

Query Match 48.7%; Score 987; DB 5; Length 1049;  
Best Local Similarity 98.3%; Pred. No. 1e-238;  
Matches 1029; Conservative 8; Mismatches 6; Indels 4; Gaps 4;

3 CGAGCGTTGAGAGACGAGAGAGAGAGAAATGGCGTCCACGGATTACAGTACCTA 62  
|||||  
7 CGAGCGTTGAGAGAGAGAGAGAGAGAGAGAAATGGCGTCCACGGATTACAGTACCTA 66  
|||||  
63 TAGCCAGCTGACGCGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 122  
|||||  
67 TAGCCAGCTGACGCGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 126  
|||||  
123 TGACAGACCAACCCAGGAGATATGGGCAACAAAGCTATGAACTATGAGACGCCACTGA 182  
|||||  
127 TGACAGACCAACCCAGGAGATATGGGCAACAAAGCTATGAACTATGAGACGCCACTGA 186  
|||||  
183 TGACAGCTATACCCAGGAGATATGGGCAACAAAGCTATGAACTATGAGACGCCACTGA 242  
|||||  
187 TGACAGCTATACCCAGGAGATATGGGCAACAAAGCTATGAACTATGAGACGCCACTGA 246  
|||||  
243 TTATGAGACAGCTCCACCTGTTATATCTATCTCAACTGCCCCCAGGAGATACAGCCAGCC 302  
|||||  
247 TTATGAGACAGCTCCACCTGTTATATCTATCTCAACTGCCCCCAGGAGATACAGCCAGCC 306  
|||||  
303 TGTCAGGGGATATGAG 362  
|||||  
307 TGTCAGGGGATATGAG 366  
|||||  
363 GGCCTCTATGAG 422  
|||||

Db 367 GGCCTCTATGAG 426  
|||  
Qy 423 GCAGCCAG 482  
|||  
Db 427 GCAGCCAG 485  
|||  
Qy 483 TCAGAGCTATGAG 542  
|||  
Db 486 TCAGAGCTATGAG 545  
|||  
Qy 543 CTACAGTTATCCAG 602  
|||  
Db 546 CTACAGTTATCCAG 605  
|||  
Qy 603 CTACAGTTATCCAG 662  
|||  
Db 606 CTACAGTTATCCAG 665  
|||  
Qy 663 TCAG 722  
|||  
Db 666 TCAG 725  
|||  
Qy 723 AACGAGTATGAG 782  
|||  
Db 726 AACGAGTATGAG 785  
|||  
Qy 783 AGCTCAAGTCAATATATAGCAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 842  
|||  
Db 786 AGCTCAAGTCAATATATAGCAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 845  
|||  
Qy 843 GAGACCAACCCAG 902  
|||  
Db 846 -GACACCCAG 904  
|||  
Qy 903 AGAGAACCCAG 962  
|||  
Db 905 AGAGAACCCAG 964  
|||  
Qy 963 TGAGAGATGAG 1022  
|||  
Db 965 TGAGAGATGAG 1023  
|||  
Qy 1023 GCGAGGTGCTTCAATATAGCCTGTGTG 1049  
|||  
Db 1024 SCGAGGTGCTTCAATATAGCCTGTGTG 1049  
|||

RESULT 15  
BX353313  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLES  
JOURNAL  
COMMENT

BX353313 1008 bp mRNA linear EST 08-APR-2004  
BX353313 Homo sapiens NEUROBLASTOMA COT 25-NORMALIZED Homo sapiens  
cDNA clone CS0D004YB22 5-PRIME, mRNA sequence.  
BX353313  
BX353313.2 GI:46291060  
EST.  
Homo sapiens (human)  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 1008)  
L1.W.B., Gruber,C., Jessee,J. and Polyes,D.  
Full-length cDNA libraries and normalization  
unpublished (2001)  
On May 5, 2003 this sequence version replaced gi:30371752.  
Contact: Genoscope  
Genoscope - Centre National de Sequencage  
2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE  
Email: segre@genoscope.cns.fr, Web : www.genoscope.cns.fr  
1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime  
end enriched, double-strand cDNA was digested with Not I and EcoR V  
sites of the pCMVSPORT 6 vector. Library was normalized. Library  
was constructed by Life Technologies, a division of Invitrogen. This  
sequence belongs to sequence cluster





GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 20, 2005, 00:41:37 ; Search time 4248.83 Seconds

(Without alignments)  
11541.209 Million cell updates/sec

Title: US-10-791-017A-1\_COPY\_1000\_2011

Perfect score: 1012

Sequence: 1 aggaacgcggtggaatcgga.....agcgacagatcgccctac 1012

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing filter 45 summaries

Database : GenBank1:\*  
1: gb\_ba:\*  
2: gb\_ptg:\*  
3: gb\_in:\*  
4: gb\_cm:\*  
5: gb\_ov:\*  
6: gb\_pat:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pr:\*  
10: gb\_ro:\*  
11: gb\_atg:\*  
12: gb\_by:\*  
13: gb\_un:\*  
14: gb\_vl:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1012	100.0	1971	12	BT007796
2	1012	100.0	1988	6	AX714390
3	1012	100.0	1988	6	AK056681
4	1012	100.0	2182	9	BC004817
5	1012	100.0	2390	6	CQ867364
6	1010.4	99.8	2371	6	A36460
7	1010.4	99.8	2371	6	AR080100
8	1010.4	99.8	2390	6	AK411125
9	1010.4	99.8	2390	6	HSEW5
10	1008.8	99.7	2026	6	CQ850483
11	1008.8	99.7	2026	6	AK127624
12	1001	98.9	2326	6	CQ730352
13	996	98.4	2164	9	AK056309
14	996	98.4	2189	9	CR456490
15	996	98.4	2364	9	BC011048
16	993.4	98.2	7293	9	BSM808920
17	955.4	94.4	2473	9	AK026270
18	911.2	90.0	193559	2	AC137500
19					AC137500 Homo sapi

20	908	89.7	155815	2	AC018774	AC018774 Homo sapi
21	908	89.7	180718	9	AL596087	AL596087 Human DNA
22	908	89.7	182501	2	AC011221	AC011221 Homo sapi
23	895.2	88.5	2189	10	BC068226	BC068226 Mus muscu
24	879.2	86.9	2188	6	AX305537	AX305537 Sequence
25	879.2	86.9	2188	10	MMEMS	X79233 M.musculus
26	879.2	86.9	2396	5	BC083960	BC083960 Xenopus 1
27	860	85.0	1783	6	CQ721057	CQ721057 Sequence
28	817.6	80.8	132906	2	AC109802	AC109802 Canis fam
29	817.6	80.8	163104	2	AC110672	AC110672 Canis fam
30	779	77.0	221444	2	AC113313	AC113313 Mus muscu
31	779	77.0	247757	2	AC121282	AC121282 Mus muscu
32	750	74.1	226400	2	AC128482	AC128482 Rattus no
33	750	74.1	225241	2	AC098231	AC098231 Rattus no
34	750	74.1	263925	2	AC106522	AC106522 Rattus no
35	653.6	64.6	1462	6	CQ434162	CQ434162 Sequence
36	595.6	58.9	2440	5	AJ719366	AJ719366 Gallus ga
37	562.8	55.6	220708	2	AC150903	AC150903 Mus muscu
38	562.8	55.6	249523	2	AC079537	AC079537 Mus muscu
39	559.6	55.3	274656	2	AC079538	AC079538 Mus muscu
40	559.6	55.3	274656	2	AC079538	AC079538 Mus muscu
41	552.8	54.6	257383	2	AC079494	AC079494 Mus muscu
42	524.6	51.8	270337	2	AC099212	AC099212 Rattus no
43	513.8	50.8	110000	2	AC149218_2	Continuation (3 of
44	510.6	50.5	160334	2	AC079549	AC079549 Mus muscu
45	508	50.2	221924	2	AC105889	AC105889 Rattus no

## ALIGNMENTS

RESULT 1	BT007796	1971 bp	mrna	linear	SYN 13-MAY-2003
LOCUS	Synthetic construct Homo sapiens Ewing sarcoma breakpoint region 1				
DEFINITION	BT007796				
ACCESSION	BT007796.1	GI:30584430			
VERSION	BT007796.1	GI:30584430			
KEYWORDS	Full cDNA.				
SOURCE	Synthetic construct				
ORGANISM	Synthetic construct				
REFERENCE	1 (bases 1 to 1971)				
AUTHORS	Kainine,N., Chen,X., Rolfs,A., Halleck,A., Hines,L., Eisenstein,S., Koundinya,M., Raphael,J., Moreira,D., Kelley,T., Labaer,J., Lin,Y., Phelan,M. and Farmer,A.				
TITLE	Cloning of human full-length cDNAs in BD Creator(TM) System Donor				
JOURNAL	Unpublished				
REFERENCE	2 (bases 1 to 1971)				
AUTHORS	Kainine,N., Chen,X., Rolfs,A., Halleck,A., Hines,L., Eisenstein,S., Koundinya,M., Raphael,J., Moreira,D., Kelley,T., Labaer,J., Lin,Y., Phelan,M. and Farmer,A.				
TITLE	Direct Submission				
JOURNAL	Submitted (13-MAY-2003) BD Biosciences Clontech, 1020 East Meadow				
COMMENT	Circle, Palo Alto, CA 94303, USA				
	This cDNA clone is a part of a collection of human full length				
	expression clones generated by BD Biosciences Clontech and the				
	Harvard Institute of Proteomics. Each cDNA has been cloned in two				
	forms: with and without stop-codon (to allow fusion with C-terminal				
	tag). The cDNA has been directionally cloned using BD In-Fusion(TM)				
	cloning system between the SalI and HindIII sites of the pDONR-DUAL				
	vector. Additional sequences in the clone: 'ACC' after SalI site				
	and before 'ATG' to provide Kozak consensus sequence; 'GG' after				
	last codon and before HindIII site to maintain reading frame.				
	clone distribution: http://bioinfo.clontech.com/orfclones.				
FEATURES	Location/Qualifiers				
Source	1..1971				
	/organism="synthetic construct"				
	/mol_type="mRNA"				
	/db_xref="taxon:32630"				
	/clone="GH00297L1.0"				
	/clone_1lb="BD Creator(TM) CDS Library derived from MGC				



Db	1040	CATCTACCTGGACAAGAAAGAAAGAAAGGCCAAAGGCGATGCCACAGTGTCTCTATGAAAGA	1099
Oy	301	CCCAACCCACTGCGCAAGGCGTGCCTGTGGATGTGTTTGAATGGGAAAGATTTTCAAGGAGCA	360
Db	1100	CCCAACCCACTGCGCAAGGCGTGCCTGTGGAAATGTTTATGGGAAAGATTTTCAAGGAGCA	1159
Oy	361	ACTTAAAGTCTCCCTTGCTCGGAAAGGCTCCATGAAACAGTATGCGGGGGTGGTCTGCC	420
Db	1160	ACTTAAAGTCTCTCTTGTCTCGGAAAGGCTCCATGAAACAGTATGCGGGGGTGGTCTGCC	1219
Oy	421	AACCCGTGAGGGCAGAGGCAATGCCCAACACTCCGTGGAGGTCCAGAGGCGCCAGAGAG	480
Db	1220	AACCCGTGAGGGCAGAGGCAATGCCCAACACTCCGTGGAGGTCCAGAGGCGCCAGAGAG	1279
Oy	481	TCCTGGGGGACCACTATGGGTCCGATGGGAGGCGGTGGAGAGATGAGAGAGCTTCCCTCC	540
Db	1280	TCCTGGGGGAGCCCAATGGGTCCGATGGGAGGCGGTGGAGAGATGAGAGAGCTTCCCTCC	1339
Oy	541	AAGAGGACCCCGGGGGTTCCGAGGGGAAACCCCTCTGGAGAGAGGAAACGTCCAGACCCGAGC	600
Db	1340	AAGAGGACCCCGGGGGTTCCGAGGGGAAACCCCTCTGGAGAGAGGAAACGTCCAGACCCGAGC	1399
Oy	601	TGAGACTGCGCAGTGTCCCAATCCGGGTTGTGGAAACAGAACTTCGCTGGAGAACAGA	660
Db	1400	TGAGACTGCGCAGTGTCCCAATCCGGGTTGTGGAAACAGAACTTCGCTGGAGAACAGA	1459
Oy	661	GTGCACAACAGTGTAAAGGCCCCCAAGCCTGAAGGCTTCTCCCGCACCTTTTCGGCCCC	720
Db	1460	GTGCACAACAGTGTAAAGGCCCCCAAGCCTGAAGGCTTCTCCCGCACCTTTTCGGCCCC	1519
Oy	721	GGGTGTGTATGTGTGCACAGGTGGGCTCTGTGTGCATGTGGGGGAGGAAGAAGTGGCTTAT	780
Db	1520	GGGTGTGTATGTGTGCACAGGTGGGCTCTGTGTGCATGTGGGGGAGGAAGAAGTGGCTTAT	1579
Oy	781	GGATCGTGTGTGGTCCCGGTGGGAATGTTCAAGAGTGGCGGTGTGTGAGACAGAGGTGGCTT	840
Db	1580	GGATCGTGTGTGGTCCCGGTGGGAATGTTCAAGAGTGGCGGTGTGTGAGACAGAGGTGGCTT	1639
Oy	841	CCGTGTGTGCGCGGGGACATGAAACCAAGGTGGCTTTGTGTGAGAGAAACAGAGTGGCTTGG	900
Db	1640	CCGTGTGTGCGCGGGGACATGAAACCAAGGTGGCTTTGTGTGAGAGAAACAGAGTGGCTTGG	1699
Oy	901	GGGGCCCCCTTGGACCTTTGATGTGAACAATGGGAGGAAAGAAAGAGGACGTGGAGGACC	960
Db	1700	GGGGCCCCCTTGGACCTTTGATGTGAACAATGGGAGGAAAGAAAGAGGACGTGGAGGACC	1759
Oy	961	TGGAATAAATGTAAAGGCGAGCACCGTCAGGAGGCGAGATGTGGGCTTAC	1012
Db	1760	TGGAATAAATGTAAAGGCGAGCACCGTCAGGAGGCGAGATGTGGGCTTAC	1811

RESULT 3	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS
AK056681		1986 bp mRNA linear PRI 30-JAN-2004	AK056681						
		Homo sapiens cDNA FLJ32119 f1s, clone FBLM1000034, highly similar to RNA-BINDING PROTEIN EWS.	AK056681						
		AK056681.1 GI:16552152							
		oligo capping; f1s (full insert sequence).							
		Homo sapiens (human)							
		Homo sapiens							
		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.							
1									
		Ota, T., Suzuki, Y., Nishikawa, T., Otsuki, T., Sugiyama, T., Irie, R., Wakamatsu, A., Hayashi, K., Sato, H., Negai, K., Kimura, K., Makita, H., Sekine, M., Odayashi, M., Nishi, T., Shibahara, T., Tanaka, T., Ishii, S., Yamamoto, J., Saito, K., Kawai, Y., Isono, Y., Nakamura, Y., Nagahara, K., Murakami, K., Yasuda, T., Iwawaghi, T., Magatsuna, M., Shirotsuki, A., Sudo, H., Hoesli, T., Kaku, Y., Kodaira, H., Kondo, H., Sugawara, M., Takahashi, M., Kanda, K., Yokoi, T., Furuya, T., Kikkawa, E., Omura, Y., Abe, K., Kamihara, K., Katsuta, N., Sato, K., Takikawa, M., Yamazaki, T., Nishimura, K., Ishibashi, T., Yamashita, H.,							

Matakawa, K., Fujimori, K., Tanai, H., Kinata, M., Watanabe, M.,  
 Hiraoka, S., Chiba, Y., Ishida, S., Ono, Y., Takiguchi, S., Watanabe, S.,  
 Yosida, M., Hotuba, T., Kusano, J., Kanehori, K., Takahashi, Fujii, A.,  
 Hara, H., Tanase, T., Nomura, Y., Togiya, S., Komai, F., Hara, R.,  
 Takeuchi, K., Arita, M., Imose, N., Mutsaers, K., Yuki, H., Oshima, A.,  
 Sasaki, N., Aotsuka, S., Yoshikawa, Y., Matsunawa, H., Ichihara, T.,  
 Shiohata, N., Sano, S., Moriya, S., Momiyama, H., Satoh, N., Takami, S.,  
 Teasahima, Y., Suzuki, O., Nakagawa, S., Senoh, A., Mizoguchi, H.,  
 Goto, Y., Shimizu, F., Wakebe, H., Hishigaki, H., Watanabe, T.,  
 Sugiyama, A., Takemoto, M., Kawakami, B., Yamazaki, M., Watanabe, K.,  
 Kumagai, A., Itakura, S., Fukuzumi, Y., Fujimori, Y., Koniyama, M.,  
 Tashiro, H., Tanigami, A., Fujiwara, T., Ono, T., Yanada, K., Fujii, Y.,  
 Ozaki, K., Hira, M., Ohmori, Y., Kawabata, A., Hiki, T., Kobatake, N.,  
 Inagaki, H., Ikema, Y., Okamoto, S., Ohtani, R., Kawakami, T.,  
 Noguchi, S., Itoh, T., Shigeta, K., Senba, T., Matsunuma, K.,  
 Nakajima, Y., Mizuno, T., Morinaga, M., Sasaki, M., Togashi, T.,  
 Oyama, M., Hata, H., Watanabe, M., Komatsu, T., Mizushima-Sugano, J.,  
 Satoh, T., Shirai, Y., Takahashi, Y., Nakagawa, K., Okumura, K.,  
 Nagase, T., Nomura, N., Kikuchi, H., Masuo, Y., Yamashita, R.,  
 Nakai, K., Yada, T., Nakamura, Y., Ohara, O., Isogai, T. and Sugano, S.  
 Complete sequencing and characterization of 21,243 full-length  
 human cDNAs  
*Nat. Genet.* 36 (1), 40-45 (2004)  
 14702039  
 2  
 Iehibashi, T., Kanehori, K., Yosida, M., Watanabe, S., Ishida, S.,

JOURNAL TITLE	REFERENCE AUTHORS TITLE JOURNAL	COMMENT
NEDO human cDNA sequencing project Unpublished	3 (bases 1 to 1986) Isogai, T., Otsuki, T. and Sugiyama, T. Direct Submission Submitted (24-OCT-2001)	Takeao Isogai, Helix Research Institute, Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0612, Japan (E-mail: genomics@hri.co.jp, Tel: 81-438-52-3975, Fax: 81-438-52-3986) NEDO human cDNA sequencing project supported by Ministry of Economy, Trade and Industry of Japan: cDNA full insert sequencing: Research Association for Biotechnology (RAB); cDNA library construction: Helix Research Institute (HRI) (supported by Japan Key Technology Center etc.); 5'- and 3'-end one pass sequencing: RAB, HRI, and Biotechnology Center, National Institute of Technology and Evaluation; clone selection for full insert sequencing: RAB and HRI.

```

FEATURES             source
location/Qualifiers
1..1988
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="PEBLM1000034"
/cell_type="peripheral blood mononuclear cells
(HPBMC593)"
/clone_11b="PEBLM1"
/note="cloning vector: PMB18FL3-primary culture,
peripheral blood mononuclear cells"
12..1814
/note="unnamed protein product"
/codon_start=1
/protein_id="BAB71252.1"
/db_xref="GI:16552153"
/translation="MASTDYSTYSQAQAQGGYAVTAQTPQGYAACTQAGVGGDSYGYT
GGTDDSYTOACTTAYTGOTAVATSGAAGGATGTTTPAPQASQPVQAGVGTGAADTTT
AATTQQAAYAAAGATGAPAYAGQQAATAPTSYSTSPQSPSYDSSAFOONTGCG
PSYGGQGSYGQDSSYGCQDPTSPYPTQGYSAQPSQYSGQSSSYGQSSAFRODHPS
MGVYGGSGGSGFSGFENRSMSPGPDNRKRGKGFDFRGMSSKRGKGGGNGSGKGERGG
FNPFGPMDEGPDLDJGPPVDPEDSDNSAIYVQGLNDSVTLDDLADFFKQCGAVKNN
KRGKQPMIHLYLDKGTGKPGADATVSEDPPTKAAVEWMDGQDPOSSKTKVSLARKK

```

## ORIGIN

PMNSMRGCLPPREGRMPPLRAGSGPGGPGGPGMGRGGRGDRGSGPPRGRGR  
GNSGGGANNVHRAGDMOCENPGGCGNPNRRTCNCKAPKPEGRPPPPRGGDRG  
RGGPGGRGGRGMDRGSGPMFRGGRGDRGGRGMDRGGRGGRGGRGSGP  
GPIBQWGRGRRGRGPGKMDKGBHRQERDRPY"

Query Match 100.0%; Score 1012; DB 9; Length 1988;  
Best Local Similarity 100.0%; Pred. No. 6.9e-197;  
Matches 1012; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGGACGGGTGATGAGGCGCTGAGAGCGAGTGGCTTCAATAGCTGTGAGC 60  
DB AGGACGGGTGATGAGGCGCTGAGAGCGAGTGGCTTCAATAGCTGTGAGC 859

QY 61 CATGATGAAAGACAGATCTTGAATCTAGGCCCTGTGATTCAGATGAATCTTGA 120  
DB CATGATGAAAGACAGATCTTGAATCTAGGCCCTGTGATTCAGATGAATCTTGA 919.

QY 121 CAACAGTGCATTTATATACAGGATTAATAAGACAGTGAATCTAGATGATGCGCAGA 180  
DB CAACAGTGCATTTATATACAGGATTAATAAGACAGTGAATCTAGATGATGCGCAGA 979

QY 181 CTCTTTAAGCAGTGTGGGGTTGTTAAGATGAACAAGAACTGGGCAACCATGATCA 240  
DB CTCTTTAAGCAGTGTGGGGTTGTTAAGATGAACAAGAACTGGGCAACCATGATCA 1039

QY 241 CATCTACCTGAGCAAGAAACAGAAACCCAAAGCGATGCCAGTGTCTTATGAAGA 300  
DB CATCTACCTGAGCAAGAAACAGAAACCCAAAGCGATGCCAGTGTCTTATGAAGA 1099

QY 301 CCCACCCACTGCCAAGGCTGCGGATGATGATGATGATGATGATGATGATGATGATGAT 360  
DB CCCACCCACTGCCAAGGCTGCGGATGATGATGATGATGATGATGATGATGATGATGAT 1100

QY 361 ACTTAAAGTCTCCCTTCTCGGAGAAAGCTTCAATGAACATGATGCGGGTGTCTGCC 420  
DB ACTTAAAGTCTCCCTTCTCGGAGAAAGCTTCAATGAACATGATGCGGGTGTCTGCC 1160

QY 421 ACCCGTGAAGGCGAGGCGATGCCAATCTCGTGAAGTGTCAAGAGGCCCAAGAGG 480  
DB ACCCGTGAAGGCGAGGCGATGCCAATCTCGTGAAGTGTCAAGAGGCCCAAGAGG 1220

QY 481 TCCGTGGGGGAGCCATGGGTTGCGATGGAGGCGGTGAGAGATGAGAGAGGCTTCCCTCC 540  
DB TCCGTGGGGGAGCCATGGGTTGCGATGGAGGCGGTGAGAGATGAGAGAGGCTTCCCTCC 1280

QY 541 AAGAGAACCCCGGGGTTCCCGAGGGAACCCCTCTGAGAGAGGAACCTCCAGCACCGAGC 600  
DB AAGAGAACCCCGGGGTTCCCGAGGGAACCCCTCTGAGAGAGGAACCTCCAGCACCGAGC 1340

QY 601 TGAAGACTGGGAGTGTCCCAATCCGGGTTGTGAGAAACCAAACTTCCCTGAGAAACAGA 660  
DB TGAAGACTGGGAGTGTCCCAATCCGGGTTGTGAGAAACCAAACTTCCCTGAGAAACAGA 1400

QY 661 GTGCAACCAAGTGAAGGCGCCAAAGGCTGAAAGGCTTCCCGCCACCTTTCCGCCCCC 720  
DB GTGCAACCAAGTGAAGGCGCCAAAGGCTGAAAGGCTTCCCGCCACCTTTCCGCCCCC 1460

QY 721 GGGTGTGATCTGTGGCAGAGGTGCGCTGTGTCATGCGGGAGGAAGAGGTGCGCTCAT 780  
DB GGGTGTGATCTGTGGCAGAGGTGCGCTGTGTCATGCGGGAGGAAGAGGTGCGCTCAT 1520

QY 781 GGATCGTGTGTGTCCCGGTGAAATGTTCAAGGTGCGCTGTGTGAGAACAGAGTGGCTT 840  
DB GGATCGTGTGTGTCCCGGTGAAATGTTCAAGGTGCGCTGTGTGAGAACAGAGTGGCTT 1580

QY 841 CCGTGTGTGCGCGGGGCAATGACCGAGGTGCTTGTGTGAGAGAAAGAGGTGCGCTTGG 900  
DB CCGTGTGTGCGCGGGGCAATGACCGAGGTGCTTGTGTGAGAGAAAGAGGTGCGCTTGG 1640

QY 901 GGGGCCCCCTTGAACCTTTGATGAAACAGATGGAGAGAAAGAGAGGACTGAGAGACC 960  
DB GGGGCCCCCTTGAACCTTTGATGAAACAGATGGAGAGAAAGAGAGGACTGAGAGACC 1700

QY 961 TGGAAAAATGATTAAGGCGAGCACCGCTGAGAGGCGCAGAGATCGGCGCTTAC 1012  
DB TGGAAAAATGATTAAGGCGAGCACCGCTGAGAGGCGCAGAGATCGGCGCTTAC 1811

## RESULT 4

BC004817

## LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

1 (bases 1 to 2182)

Klausner, R.D., Collins, P.S., Wagner, L., Shemen, C.M., Schuler, G.D.,

Altschul, S.F., Zeeberg, B., Buecaw, K.H., Schaefer, C.P., Bhat, N.K.,

Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, D., Hsieh, P.,

Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,

Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L.,

Scheltz, T.E., Brownstein, M.J., Uedlin, T.B., Toshitoki, S.,

Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J.,

Abramson, R.D., Mullany, S.J., Bosak, S.A., McEwan, P.D.,

McKernan, K.J., Malek, J.A., Gunaratne, P.H., Gay, L.J., Hulyk, S.W.,

Morley, R.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W.,

Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,

Fahey, J., Helton, B., Ketterman, M., Madan, A., Young, A.C., Shevchenko, Y.,

Sanchez, A., Whiting, M., Madan, A., Touchman, J.W., Green, E.D.,

Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,

Butterfield, Y.S., Krzywinski, M.I., Skalske, U., Smalton, D.E.,

Schermer, A., Schein, J.E., Jones, S.J., and Marra, M.A.

Generation and initial analysis of more than 15,000 full-length

human and mouse cDNA sequences

Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)

12477932

2 (bases 1 to 2182)

Strausberg, R.

Direct Submission

Submitted (21-MAR-2001) National Institutes of Health, Mammalian

Gene Collection (MGC), Cancer Genomics Offices, National Cancer

Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2550,

USA

NIH-MGC Project URL: <http://mgc.ncl.nih.gov>

Contact: MGC help desk

Email: [cgabs-remail.nih.gov](mailto:cgabs-remail.nih.gov)

Tissue Procurement: ATCC

CDNA Library Preparation: Life Technologies, Inc.

DNA Sequencing Arrayed by: The I.M.A.G.E. Consortium (LNL)

Sequencing Center

Center code: BCM-HGSC

Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>Contact: [amg@bcm.tmc.edu](mailto:amg@bcm.tmc.edu)

Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Louisse, H.,

Kowis, C.R., Sneed, A.J., Martin, R.G., Muzny, D.M., Nanavati,

A.N., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found

through the I.M.A.G.E. Consortium/LNL at: <http://image.lnl.gov>

Series: IRAC Plate: 3 Row: 1 Column: 6

This clone was selected for full length sequencing because it

passed the following selection criteria: Hexamer frequency ORF

analysis.

Location/Qualifiers

1. 2182

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"  
 /clone="MGC:5432.1MGC:3449145"  
 /issue\_type="Placenta, choriocarcinoma"  
 /clone\_lib="NH\_MGC\_10"  
 /lab\_host="DH10B"  
 /note="Vector: pCMV-Sport6"  
 1. 2182  
 /gene="BMSR1"  
 /note="synonym: EMS"  
 /db\_xref="LOCUSID:2130"  
 /db\_xref="MIM:133450"  
 20. 1990  
 /gene="BMSR1"  
 /codon\_start=1  
 /product="Fbwing sarcoma breakpoint region 1, isoform EMS"  
 /protein\_id="AAH04817.1"  
 /db\_xref="GI:13435963"  
 /db\_xref="LOCUSID:2130"  
 /db\_xref="MIM:133450"  
 /translation="MSTSTYSQAAAOOGYSAYTACPGVAGTQTAAYGQSGYVY  
 GQPTDSTAOQTATATGTVGATVYSGGPTEGTTPTAAOASVQGVGGYADITTT  
 ALVTITQASVAASGAVETQDPAVAGQAPPTATSTPTDPTSYNDSTQSTNYYGQPS  
 BSLTGQSTSYGQDSSTPGQPPSTPTPTQGSTSQAPSTQSDSSSYGQDSFRDHDHSSNG  
 SYGQDSSTYGQDSSTPGQPPSTPTPTQGSTSQAPSTQSDSSSYGQDSFRDHDHSSNG  
 VYGGQSGGDFDLEGRVMSQSDNRGRGQFPRGKWSRGGRGSGMGSGAGEGRGSG  
 KPGPMEGEGDLIDGPVDEDDSDNSALYVGLDLSVTLIDLDPFKQCGVYQNNRKR  
 TGSMRIYLIDKEGRKQGYATSEYDPTAKAAVEMFGKQFQSKLSTLARKPP  
 NNNRRGGLPPEGRGAMPPLRGGRGGPGGPGMGRGGGRGGRGGRGGRGRGRGR  
 PGSGGQVHAAGMQCNCPCGQNNANNTTECNCKAPRBEFLPPPPPPCGDGRGCG  
 GCGNRGGRGGLNDRGPGGEMFRHGGNGRGRGRGRGRGRGRGRGRGRGGPGGPGP  
 MMEWGRRGRGRGGPGMGMDGHRDGRRRPPY

## ORIGIN

Query Match	Similarity	100.0%	Score 1012;	DB 9;	Length 2182;
Best Local	Similarity	100.0%	Pred. No. 6.8e-157;		
Matches 1012;	Conservative	0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	AGACGCGGTGGATGGCGAGCGCTGGAGAGCGAGGTGGCTTCATATAGCTGGTGGACC	60		
Db	976	AGGACGCGGTGGAAATGGGCGAGCGCTGGAGAGAGGTGGCTTCATATAGCTGGTGGACC	1035		
QY	61	CATGATGAAGGACCAATCTTGAATCTTAGGCCCACTGTAGATCCAGATGAAGACTCTGA	120		
Db	1036	CATGATGAAGGACCAAGATCTTGAATCTTAGGCCCACTGTAGATCCAGATGAAGACTCTGA	1095		
QY	121	CAACAGTGCATTTTATGTATCAAGAGTTAAATGACAGGTGACTCTAGATGATCTAGGCGA	180		
Db	1096	CAACAGTGCATTTTATGTATCAAGAGTTAAATGACAGGTGACTCTAGATGATCTAGGCGA	1155		
QY	181	CTTCTTTAAGCAGTGTGGGGTTGTAAATGAAACAAGAACTGGGCAACCATGATCCA	240		
Db	1156	CTTCTTTAAGCAGTGTGGGGTTGTAAATGAAACAAGAACTGGGCAACCATGATCCA	1215		
QY	241	CATCTACCTTGGACAGAGAAACAGAAAGCCCAAGCGATGCCACAGTGTCTTATGAAGA	300		
Db	1216	CATCTACCTTGGACAGAGAAACAGAAAGCCCAAGCGATGCCACAGTGTCTTATGAAGA	1275		
QY	301	CCCAACCCACTGCGAAGGCGTCCCGTGGAAATGTTTGAATGGAAAGATTTTCAAGGGACAA	360		
Db	1276	CCCAACCCACTGCGAAGGCGTCCCGTGGAAATGTTTGAATGGAAAGATTTTCAAGGGACAA	1335		
QY	361	ACTTAAATCTCCCTTGTCTCGAAGAAAGCTTCAATGAACAGTATGCGGGGTGGTCTGCC	420		
Db	1336	ACTTAAATCTCCCTTGTCTCGAAGAAAGCTTCAATGAACAGTATGCGGGGTGGTCTGCC	1395		
QY	421	ACCCGTTAGGGGACAGAGGCGATGCCACCACTCCGTTGAAGGTCCAGAGGCCCCAGAGG	480		
Db	1396	ACCCGTTAGGGGACAGAGGCGATGCCACCACTCCGTTGAAGGTCCAGAGGCCCCAGAGG	1455		
QY	481	TCCTGGGGGGACCCATGGGTGCCATGGGAGGCGCTGGAGAGAAATGAAGAGGCTTCCCTCC	540		
Db	1456	TCCTGGGGGGACCCATGGGTGCCATGGGAGGCGCTGGAGAGAAATGAAGAGGCTTCCCTCC	1515		

OY	541	AAGAGGACCCCGGGGTTTCCGAGGGAAACCCCTCTGGAGAGAGAAACGTCCAGCACCGAGC	600
Db	1516	AAGAGACCCCGGGGTTTCCGAGGGAAACCCCTCTGGAGAGAGAAACGTCCAGCACCGAGC	1575
OY	601	TGAGAGCTGGCAGTGTGCCAATCCGAGGTGTGGAAACAGAACTTGCCTTGAGAAAGAG	660
Db	1576	TGAGAGACTGGCAGTGTGCCAATCCGAGGTGTGGAAACAGAACTTGCCTTGAGAAACAGAG	1635
OY	661	GTGCACACAGTGTAAAGGCCCCAAAGCCTGAAAGGTTTCTCCCGCACCTTTCCGCCCCC	720
Db	1636	GTGCACACAGTGTAAAGGCCCCAAAGCCTGAAAGGTTTCTCCCGCACCTTTCCGCCCCC	1695
OY	721	GGGTGTGATCTGTGGCAGAGGTGGCCCTGTGTGCATCGGGGAGAGAAAGGTGGCTCAT	780
Db	1696	GGGTGTGATCTGTGGCAGAGGTGGCCCTGTGTGCATCGGGGAGAGAAAGGTGGCTCAT	1755
OY	781	GGATCTGTGTGTCTCCCGGTGGATGTTCACAGAGTGGCCGTGTGTGAGACAGAGGTGCTT	840
Db	1756	GGATCTGTGTGTCTCCCGGTGGATGTTCACAGAGTGGCCGTGTGTGAGACAGAGGTGCTT	1815
OY	841	CCGTGTGTGGCCCGGGGACATGGACCGAGGTGGCTTTGTGTGAGAGAAAGAGAGGTGGCCCTGG	900
Db	1816	CCGTGTGTGGCCCGGGGACATGGACCGAGGTGGCTTTGTGTGAGAGAAAGAGAGGTGGCCCTGG	1875
OY	901	GGGGCCCCCTTGAGACTTTGTATGGAAACAGATGGAGAGAAAGAGAGAGACGTGGAGACC	960
Db	1876	GGGGCCCCCTTGAGACTTTGTATGGAAACAGATGGAGAGAAAGAGAGAGACGTGGAGACC	1935
OY	961	TGAAAAAATGATTTAAAGCGAGCACCGTCAGAGACCGAGAGATTCGGCCCTTAC	1012
Db	1936	TGAAAAAATGATTTAAAGCGAGCACCGTCAGAGACCGAGAGATTCGGCCCTTAC	1987

RESULT 5	LOCUS	DEFINITION	ACCESSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS	JOURNAL	FEATURES
LOCUS	COB67364	Sequence 1 from Patent EPI455190.	COB67364	GI:51997589		Homo sapiens (human)	Obendorf, M. and Wolf, S.			Modulation of the activity of nuclear receptors via EMS Patent: EP 1455190-A 1 08-SEP-2004;
DEFINITION	COB67364					Homo sapiens				Location/Qualifiers
ACCESSION	COB67364.1									1. .2390
KEYWORDS										/organism="Homo sapiens"
SOURCE										/mol_type="unassigned DNA"
ORGANISM										/db_xref="taxon:9606"
REFERENCE										44. .2014
AUTHORS										/note="unnamed protein product"
JOURNAL										/codon_start=1
FEATURES										/protein_id="CAH33891.1"
										/db_xref="GI:51997590"
										/translation="MASTDYSTVYSQAQAAGQGSAYTAQPTQGYAQTTOAYAGQGSYGT
										GOPTDYSYTOAQTTATYGCATVATVSYGQPTGTTTPTAPAYGSAQPVQGYGTGAVDIT
										ATVTTQASVAAQSAVGTOAPAYVGGQAPAAATPTSPDQNKFTYETSQPOSSGTGVNQ
										PSLGVQSSSYSPVQVGSYMPQVTAAPVAPTSYSTQPTSYDQSSSQOONLYGQPS
										SYQDQSSYGGQSSSYGQQPTPSYTPQTSYGSQAASQVSYQDSSSYGGQSSFPQDIPSSNG
										VYQDQSSGFSFGPEENRSMGSPDNRGRGRGFFDQGMRSRGSRGGRGSGMGSAGRGGFTN
										KQSPGMDGPDLDLGPVDEDEDSDNSAIVTGLNDSVTLLDLADLAFKQGVVAMNKR
										TGGPMHIIYLDKETGKRGDATTSEYEDPTAKAAVEMFDGKDFQGSKLKYLARKKP
										MSNRGKPLPREERGMKPPPLRGPGGPGGPGMNRGMRGGRGGRGPGPPLPGSRGNG
										PSGGANNVQHARGDMOCNPPCGNQNQNPAMRTECNQCKAPKPEGFLPPFPFPGGDRGN
										GPGRMGRGRGLMDNRGPGGMFPFGKGGGDRGGRGRGMDNRGGFGGGRRGPGGPPGP
										LMWQGGRRGRGRGPGMGRGHEHQERDRPY"

Query Match 100.0%; Score 1012; DB 6; Length 2390;  
 Best Local Similarity 100.0%; Pred. No. 6 8e-197;  
 Matches 1012; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 AGGACCGGCTGGAATGAGGAGCGCTGGAGAGCGAGTGGCTTCAATAGCTGTGGACC 60  
 1000 AGGACCGGCTGGAATGAGGAGCGCTGGAGAGCGAGTGGCTTCAATAGCTGTGGACC 1059

61 CATGATGAGAGACCAATCTTATCTAGAGCCCACTGTATGATCCAGATGAACCTCTGA 120  
 1060 CATGATGAGAGACCAATCTTATCTAGAGCCCACTGTATGATCCAGATGAACCTCTGA 1119

121 CAACAGTGAATTTATGATCAAGATTAATGACAGTGTACTCTAGATGATCTGGAGA 180  
 1120 CAACAGTGAATTTATGATCAAGATTAATGACAGTGTACTCTAGATGATCTGGAGA 1179

181 CTTCTTTAGCAGTGTGGGTTGTTAATGATGAACAGAGAACTGGGCAACCATGATCCA 240  
 1180 CTTCTTTAGCAGTGTGGGTTGTTAATGATGAACAGAGAACTGGGCAACCATGATCCA 1239

241 CATCTACTGGAACAAGAAACAGAAAGCCCAAGGAGATGCCACAGTGTCTTATGAAGA 300  
 1240 CATCTACTGGAACAAGAAACAGAAAGCCCAAGGAGATGCCACAGTGTCTTATGAAGA 1299

301 CCCACCACTGCGCAAGGCTCGGTGATGATGATGATGAGAAAGATTTTCAAGGAGCAA 360  
 1300 CCCACCACTGCGCAAGGCTCGGTGATGATGATGATGAGAAAGATTTTCAAGGAGCAA 1359

361 ACTTAAAGTCTCCCTTCTCTGGAGAAAGCTTCAATGAACATATGCGGGGTGCTGCC 420  
 1360 ACTTAAAGTCTCCCTTCTCTGGAGAAAGCTTCAATGAACATATGCGGGGTGCTGCC 1419

421 ACCCTGTGAGGAGGAGGAGGATGCCACCACTCTCTGTGAGTGTCCAGAGAGCCAGAGG 480  
 1420 ACCCTGTGAGGAGGAGGAGGATGCCACCACTCTCTGTGAGTGTCCAGAGAGG 1479

481 TCCTGGGGGAGCCCATGAGGTGATGAGAGGCGGTGAGAGATGAGAGAGGCTTCCCTCC 540  
 1480 TCCTGGGGGAGCCCATGAGGTGATGAGAGGCGGTGAGAGATGAGAGAGGCTTCCCTCC 1539

541 AAGAGACCCCGGGGTTCCCGAGGAAACCCCTCTGAGAGAGAAACCTCCAGACCGAGC 600  
 1540 AAGAGACCCCGGGGTTCCCGAGGAAACCCCTCTGAGAGAGAAACCTCCAGACCGAGC 1599

601 TGGAGATGAGAGTGTCCCAATCCGGGTGTGAGAAACCAAGAACTTCCCTGTGAGAAAGA 660  
 1600 TGGAGATGAGAGTGTCCCAATCCGGGTGTGAGAAACCAAGAACTTCCCTGTGAGAAAGA 1659

661 GTGCAACCAATGTAAAGCCCAAGAGCTTGAAGGCTTCTCCCGCACCTTCCGCCCCC 720  
 1660 GTGCAACCAATGTAAAGCCCAAGAGCTTGAAGGCTTCTCCCGCACCTTCCGCCCCC 1719

721 GGGTGTGATGTGTGCAAGAGTGGCTGTGTGCAATGCGGGAGAGAAAGAGTGGCTCAT 780  
 1720 GGGTGTGATGTGTGCAAGAGTGGCTGTGTGCAATGCGGGAGAGAAAGAGTGGCTCAT 1779

781 GGAATGTGTGTGTCCCGTGTGAATGTTCAAGAGTGTGCGGTGTGAGAAAGAGTGTGCTT 840  
 1780 GGAATGTGTGTGTCCCGTGTGAATGTTCAAGAGTGTGCGGTGTGAGAAAGAGTGTGCTT 1839

841 CCGTGTGTGTGTCCCGTGTGAATGTTCAAGAGTGTGCGGTGTGAGAAAGAGTGTGCTT 900  
 1840 CCGTGTGTGTGTCCCGTGTGAATGTTCAAGAGTGTGCGGTGTGAGAAAGAGTGTGCTT 1899

901 GGGGCCCCCTTGACCTTTGATGAAACAGATGAGAGAAAGAGAGAGAGCTGGAGAGCC 960  
 1900 GGGGCCCCCTTGACCTTTGATGAAACAGATGAGAGAAAGAGAGAGAGCTGGAGAGCC 1959

961 TGGAAAAATGATTAAGGCGAGCAACCTTCAGAGAGCGAGAGATGCGGCCCTTAC 1012  
 1960 TGGAAAAATGATTAAGGCGAGCAACCTTCAGAGAGCGAGAGATGCGGCCCTTAC 2011

RESULT 6  
 A36460 2371 bp DNA linear PAT 05-MAR-1997  
 LOCUS A36460  
 DEFINITION Sequence 1 from Patent WO9323549.  
 ACCESSION A36460  
 VERSION A36460.1 GI:2293778  
 KEYWORDS  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

REFERENCE  
 1 (bases 1 to 2371)  
 Aurias, A., Delattre, O., Desmaziere, C., Melot, T., Peter, M., Plougastel, B., Thomas, G. and Zucman, J.  
 NUCLEIC ACID CORRESPONDING TO A GENE OF CHROMOSOME 22 INVOLVED IN RECURRENT CHROMOSOMAL TRANSLOCATIONS ASSOCIATED WITH THE DEVELOPMENT OF CANCEROUS TUMORS  
 Patent: WO 9323549-A 1 25-NOV-1993;  
 CENTRE NAT RECH SCIENT (FR)  
 Other publication FR 2691475 931126  
 Other publication JP 8500964T 960206.  
 Location/Qualifiers  
 1..2371  
 /organism="Homo sapiens"  
 /mol\_type="unassigned DNA"  
 /db\_xref="taxon:9606"  
 /dev\_stage="FOETUS"

ORIGIN  
 Query Match 99.8%; Score 1010.4; DB 6; Length 2371;  
 Best Local Similarity 99.9%; Pred. No. 1.4e-196;  
 Matches 1011; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

1 AGGACCGGCTGGAATGAGGAGCGCTGGAGAGCGAGTGGCTTCAATAGCTGTGGACC 60  
 981 AGGACCGGCTGGAATGAGGAGCGCTGGAGAGCGAGTGGCTTCAATAGCTGTGGACC 1040

61 CATGATGAGAGACCAATCTTATCTAGAGCCCACTGTATGATCCAGATGAACCTCTGA 120  
 1041 CATGATGAGAGACCAATCTTATCTAGAGCCCACTGTATGATCCAGATGAACCTCTGA 1100

121 CAACAGTGAATTTATGATCAAGATTAATGACAGTGTACTCTAGATGATCTGGAGA 180  
 1101 CAACAGTGAATTTATGATCAAGATTAATGACAGTGTACTCTAGATGATCTGGAGA 1160

181 CTTCTTTAGCAGTGTGGGTTGTTAATGATGAACAGAGAACTGGGCAACCATGATCCA 240  
 1161 CTTCTTTAGCAGTGTGGGTTGTTAATGATGAACAGAGAACTGGGCAACCATGATCCA 1220

241 CATCTACTGGAACAAGAAACAGAAAGCCCAAGGAGATGCCACAGTGTCTTATGAAGA 300  
 1221 CATCTACTGGAACAAGAAACAGAAAGCCCAAGGAGATGCCACAGTGTCTTATGAAGA 1280

301 CCCACCACTGCGCAAGGCTCGGTGATGATGATGATGAGAAAGATTTTCAAGGAGCAA 360  
 1281 CCCACCACTGCGCAAGGCTCGGTGATGATGATGATGAGAAAGATTTTCAAGGAGCAA 1340

361 ACTTAAAGTCTCCCTTCTCTGGAGAAAGCTTCAATGAACATATGCGGGGTGCTGCC 420  
 1341 ACTTAAAGTCTCCCTTCTCTGGAGAAAGCTTCAATGAACATATGCGGGGTGCTGCC 1400

421 ACCCGTGAAGGAGAGAGATGCCACCACTCTGTGAGAGTTCAGAGAGCCAGAGAG 480  
 1401 ACCCGTGAAGGAGAGAGATGCCACCACTCTGTGAGAGTTCAGAGAGCCAGAGAG 1460

481 TCCTGGGGGAGCCCATGAGGTGCAATGAGAGGCGGTGAGAGAGATGAGAGAGCTTCCCTCC 540  
 1461 TCCTGGGGGAGCCCATGAGGTGCAATGAGAGGCGGTGAGAGAGATGAGAGAGCTTCCCTCC 1520

541 AAGAGACCCCGGGGTTCCCGAGGAAACCCCTTCTGAGAGAGAGAAACCTCCAGACCGAGC 600  
 1521 AAGAGACCCCGGGGTTCCCGAGGAAACCCCTTCTGAGAGAGAGAAACCTCCAGACCGAGC 1580



QY	601	GGAGAGCTGGGAGAGTGTCCCATCCGGGTTGTGGAAACAGAACTTCGCTGGAAGACGA	660
Db	1581	TGAGAGCTGGCAGTGTCCCAATCCGGGTGTGGAAACAGAACTTCGCTGAGAACGA	1640
QY	661	GTGCACCAAGTGTAAAGGCCCCAAAGCCTGAAGCTTCTCCCGCACCTTTCGCGCCCC	720
Db	1641	GTGCACCAAGTGTAAAGGCCCCAAAGCCTGAAGCTTCTCCCGCACCTTTCGCGCCCC	1700
QY	721	GGGTGTGTATCGTGCAGAGGTGGCCCTGTGTGACATCGCGGAGGAAAGGTGGCTTCAT	780
Db	1701	GGGTGTGTATCGTGCAGAGGTGGCCCTGTGTGACATCGCGGAGGAAAGGTGGCTTCAT	1760
QY	781	GGATCGTGTGTGTCCCGGTGGAAATTTCAAGAGTGGCCGTGTGTGAGACAAGTGTGCTT	840
Db	1761	GGATCGTGTGTGTCCCGGTGGAAATTTCAAGAGTGGCCGTGTGTGAGACAAGTGTGCTT	1820
QY	841	CCGTGTGTGGCCGGGCGATGGAACCGAGTGTGCTTGTGTGAGAGAAACAGGTGTGGCCCTGG	900
Db	1821	CCGTGTGTGGCCGGGCGATGGAACCGAGTGTGCTTGTGTGAGAGAAACAGGTGTGGCCCTGG	1880
QY	901	GGGGCCCCCTTGGACCTTTGATGGAACAATGGAGGAAGAAAGAGAGGACGTGAGAGACC	960
Db	1881	GGGGCCCCCTTGGACCTTTGATGGAACAATGGAGGAAGAAAGAGAGGACGTGAGAGACC	1940
QY	961	TGGAATAATGATTAAGCGCGAGCACCGTCAAGAGCGCAAGATCGGCCCTTAC	1012
Db	1941	TGGAATAATGATTAAGCGCGAGCACCGTCAAGAGCGCAAGATCGGCCCTTAC	1992

LOCUS	AR080100	2371 bp	DNA	linear	PAT 31-AUG-2000
DEFINITION	Sequence 1 from patent US 5968734.				
ACCESSION	AR080100				
VERSION	AR080100.1	GI:10006835			
KEYWORDS	.				
SOURCE	Unknown.				
ORGANISM	Unknown.				
REFERENCE	Unclassified.				
AUTHORS	1 (bases 1 to 2371)				
	Aurila, A., Delattre, O., Desmazes, C., Melot, T., Peter, M.,				
	Ploogastel, B., Thomas, G. and Zucman, J.				
TITLE	Nucleic acid corresponding to a gene of chromosome 22 involved in recurrent chromosomal translocations associated with the development of cancerous tumors, and nucleic acids of fusion resulting from said translocations				
JOURNAL	Patent: US 5968734-A 1 19-OCT-1999,				
FEATURES	Location/Qualifiers				

```

/mol_type="unassigned DNA"
ORIGIN

```

Query Match	99.8%	Score 1010.4	DB 6	Length 2371
Best Local Similarity	99.9%	Pred. No. 1.4e-196		
Matches 1011	Conservative 0	Mismatches 1	Indels 0	Gaps 0
QY	1	AGACGCGGTGGAATGGGCAACGCTGAGAGACCGAGGTGGCTTCATATAGCCTGGTGACC	60	
Db	981	AGGACGCGGTGGAATGGGCAACGCTGAGAGACCGAGGTGGCTTCATATAGCCTGGTGACC	1040	
QY	61	CATGATGTAAGAAGCAAGATCTTGATCTTGAGGCCCACTGTAAATCAATGAAAGCTCTGA	120	
Db	1041	CATGATGTAAGAAGCAAGATCTTGATCTTGAGGCCCTCTGTAAATCAAGTGAAGACTCTGA	1100	
QY	121	CAACAGTGCAATTTATGTACAAGATTAAATGACAGTGTGACTAGATGATCTGGCAGA	180	
Db	1101	CAACAGTGCAATTTATGTACAAGATTAAATGACAGTGTGACTAGATGATCTGGCAGA	1160	
QY	181	CTTCTTTAAGCAGGTGTGGGTTGTTAAGATGAACAAGAACTGGGCAACCCATGATCCA	240	
Db	1161	CTTCTTTAAGCAGGTGTGGGTTGTTAAGATGAACAAGAACTGGGCAACCCATGATCCA	1220	

Qy	241	CATCTACTGGACAAGAAACAGGAAGCCCAAGGGGGGCCACAGTGTCTTATGAAGA	300
Db	1221	CATTCTACTGGACAAGAAACAGGAAGCCCAAGGGGGGATGCCACAGTGTCTATGAAGA	1280
Qy	301	CCCAACCACTGCGCAGAGCTGCGCTGGATGGATTGATGGGAAAGATTTTCAAGGAGCAA	360
Db	1281	CCCAACCACTGCGCAGAGCTGCGCTGGATGGATTGATGGGAAAGATTTTCAAGGAGCAA	1340
Qy	361	ACTTAAAGTCTCCCTTGTCTGGAAAGAAAGCTCCAAATGACATATGCGGGGTGTCTGCC	420
Db	1341	ACTTAAAGTCTCCCTTGTCTGGAAAGAAAGCTCCAAATGACATATGCGGGGTGTCTGCC	1400
Qy	421	ACCCCGTAGGGGCGAAGGCAATGCCACACCACTCCGTGGAGGTCCAGAGGCCCAAGAGG	480
Db	1401	ACCCCGTAGGGGCGAAGGCAATGCCACACCACTCCGTGGAGGTCCAGAGGCCCAAGAGG	1460
Qy	481	TCCTGGGGGGAACCAATGGGTCGCATGGAGAGGCCGTGGAGAGATAGAGAGGCTTCCCTCC	540
Db	1461	TCCTGGGGGGAACCAATGGGTCGCATGGAGAGGCCGTGGAGAGATAGAGAGGCTTCCCTCC	1520
Qy	541	AAGAGAACCCCGGGGTTTCCGAGGGAAACCCCTCTGGAGAGAGAAACGTCCAGCACCGAGC	600
Db	1521	AAGAGAACCCCGGGGTTTCCGAGGGAAACCCCTCTGGAGAGAGAAACGTCCAGCACCGAGC	1580
Qy	601	TGGAGACTGGCAGTGTCCCAATCCGGGTTGTGGAAACCAAACTTCCGCTGGAGAAACAGA	660
Db	1581	TGGAGACTGGCAGTGTCCCAATCCGGGTTGTGGAAACCAAACTTCCGCTGGAGAAACAGA	1640

Position	Sequence	Position	Sequence
Db	1641 GTGCACCAAGTGAAGGCCCCAAAGCTTGAAGCTTCTCCCGCACCTTTCCGCCCC	1700	
Qy	721 GGGTGTGATCGTGGCAGAGGTGGCCCTGTGTGCATCGGGAGAAAGGTGGCTCAT	780	
Db	1701 GGGGTGTGATCGTGGCAGAGGTGGCCCTGTGTGCATCGGGAGAAAGGTGGCTCAT	1760	
Qy	781 GGATCGTGTGTCCTCCGCTGGAATGTTCAAGGTGGCCGTGTGAAACAGAGTGGCTT	840	
Db	1761 GGATCGTGTGTCCTCCGCTGGAATGTTCAAGGTGGCCGTGTGAAACAGAGTGGCTT	1820	
Qy	841 CCGTGTGTGCCCCGGGGCATGACCGAGGTGGCTTTGTGTGAGAAAGACGAGGTGGCCCTGG	900	
Db	1821 CCGTGTGTGCCCCGGGGCATGACCGAGGTGGCTTTGTGTGAGAAAGACGAGGTGGCCCTGG	1880	
Qy	901 GGGGCCCCCTTGACCTTTGATGAAACGATGGGAGAAAGAGAGACGTTGAGGAC	960	
Db	1881 GGGGCCCCCTTGACCTTTGATGAAACGATGGGAGAAAGAGAGACGTTGAGGAC	1940	
Qy	961 TGGAAAAATGATTAAGCGACGACCCGTCAGAGCCGAGAGATCGGCTTCAC	1012	
Db	1941 TGGAAAAATGATTAAGCGACGACCCGTCAGAGCCGAGAGATCGGCTTCAC	1992	

RESULT 8  
AX411125

```

LOCUS       AX411125                2390 bp          linear
DEFINITION  Sequence 3772 from Patent WO0229103.
ACCESSION   AX411125
VERSION     AX411125.1
KEYWORDS    GI:21443830

SOURCE      *
ORGANISM    Homo sapiens (human)
            Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE   1
AUTHORS    Alvares,C., Horne,D., Peres-da-Silva,S. and Vockley,J.G.
TITLE      Gene expression profiles in liver cancer
JOURNAL    Patent: WO 0229103-A 3772 11-Apr-2002;
FEATURES    (1)
            GENE LOCIC INC (US)
            Location/Qualifiers
            1..2390
            /organism="Homo sapiens"
            /mol_type="unassigned DNA"

```

## ORIGIN

/db\_xref="taxon:9606"  
/note="EMBL/Genbank Accession No. X66899"

```

Query Match      99.8%; Score 1010.4; DB 6; Length 2390;
Best Local Similarity 99.9%; Pred. No. 1,4e-196;
Matches 1011; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AGAAGCGGTGGTGAATGGGCGCGCTGGAGAGCCGAGTGGCTTCAATGAGCTGTGGAGC 60
DB 1000 AGAAGCGGTGGTGAATGGGCGCGCTGGAGAGCCGAGTGGCTTCAATGAGCTGTGGAGC 1059
QY 61 CATGATGAAGGACCGAGATCTTGATCTAGAGCCCACTGTAGATCCAGATGAAGCTCTGA 120
DB 1060 CATGATGAAGGACCGAGATCTTGATCTAGAGCCCTCTGTAGATCCAGATGAAGCTCTGA 1119
QY 121 CAACAGTGAATTTATGTAAGAGATTTAAATGACAGTGTACTCTAGATGATCTGGCAGA 180
DB 1120 CAACAGTGAATTTATGTAAGAGATTTAAATGACAGTGTACTCTAGATGATCTGGCAGA 1179
QY 181 CTTCTTTAAGCAGTGTGGGTTGTTAAGATGAAACAGAGAACTGGGCAACCCATGATCCA 240
DB 1180 CTTCTTTAAGCAGTGTGGGTTGTTAAGATGAAACAGAGAACTGGGCAACCCATGATCCA 1239
QY 241 CATCTACCTGGAACAAGGAAACAGGAAAGCCCAAGGCGATGCCAGTGTCTATGAGA 300
DB 1240 CATCTACCTGGAACAAGGAAACAGGAAAGCCCAAGGCGATGCCAGTGTCTATGAGA 1299
QY 301 CCCACCACTGCCAAGGCTGCGCTGAGATGTTGATGGAGAAATTTTCAAGGGAGCAA 360
DB 1300 CCCACCACTGCCAAGGCTGCGCTGAGATGTTGATGGAGAAATTTTCAAGGGAGCAA 1359
QY 361 ACTTAAAGTCTCTCTGCTGGAGAAAGCTTCCATGAACATGATGCGGGGTGTCTGCC 420
DB 1360 ACTTAAAGTCTCTCTGCTGGAGAAAGCTTCCATGAACATGATGCGGGGTGTCTGCC 1419
QY 421 ACCCGGAGGGGAGAGGAGATGCCACCACTCGTGAAGGTCCAGAGAGGCCAGGAGG 480
DB 1420 ACCCGGAGGGGAGAGGAGATGCCACCACTCGTGAAGGTCCAGAGAGGCCAGGAGG 1479
QY 481 TCCTGGGGGAGCCCATGGGTGCGATGGAGGCGCTGGAGAGATGAGAGAGCTTCCCTCC 540
DB 1480 TCCTGGGGGAGCCCATGGGTGCGATGGAGGCGCTGGAGAGATGAGAGAGCTTCCCTCC 1539
QY 541 AAGAGACCCCGGGGCTTCCGAGAGAAACCTCTCTGAGAGAGAAAGCTCCAGCACCGAGC 600
DB 1540 AAGAGACCCCGGGGCTTCCGAGAGAAACCTCTCTGAGAGAGAAAGCTCCAGCACCGAGC 1599
QY 601 TGGAGACTGGGAGGTGCCAATCCGGGTTGTGAGAAACAGAACTTGCCTGGAGACAGA 660
DB 1600 TGGAGACTGGGAGGTGCCAATCCGGGTTGTGAGAAACAGAACTTGCCTGGAGACAGA 1659
QY 661 GTGCACACAGTGAAGGCCCAAGACCTGAAAGCTTCTCCGCGACACCTTTCGCCCCC 720
DB 1660 GTGCACACAGTGAAGGCCCAAGACCTGAAAGCTTCTCCGCGACACCTTTCGCCCCC 1719
QY 721 GGGGTGTGATGTGGCAGAGGTGGCCTGTGTGCAATGCGGGGAGAGAGAGGTGGCTCAT 780
DB 1720 GGGGTGTGATGTGGCAGAGGTGGCCTGTGTGCAATGCGGGGAGAGAGAGGTGGCTCAT 1779
QY 781 GGATCGTGTGTGTCCTCGGTGGAATGTTCAAGAGTGGCCGTGTGAGAGACAGAGGTGGCTT 840
DB 1780 GGATCGTGTGTGTCCTCGGTGGAATGTTCAAGAGTGGCCGTGTGTGAGAGACAGAGGTGGCTT 1839
QY 841 CCGTGTGTGCGCGGGGAGATGAGACCGAGGTGCTTGTGTGAGAGAGACAGAGGTGGCTTGG 900
DB 1840 CCGTGTGTGCGCGGGGAGATGAGACCGAGGTGCTTGTGTGAGAGAGACAGAGGTGGCTTGG 1899
QY 901 GGGGCCCCCTTGAACCTTTGATGAGAACAGATGGAGAGAAAGAGAGAGAGCTGTGAGAGC 960
DB 1900 GGGGCCCCCTTGAACCTTTGATGAGAACAGATGGAGAGAAAGAGAGAGAGCTGTGAGAGC 1959
QY 961 TGGAAAAATGATTAAGCGGAGCAACCGTCAAGAGGCGAGAGATGGCCCTAC 1012

```

```

Db 1960 TGGAAAAATGATTAAGCGGAGCACCGTCAAGAGCGCAGAGATGGCCCTAC 2011

RESULT 9
HSEMS LOCUS H. sapiens EMS mRNA. 2390 bp mRNA linear PRI 28-JUN-1995
DEFINITION X66899
ACCESSION X66899.1 GI:547565
VERSION RNA binding protein.
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
Delattre, O., Zucman, J., Plougaeste, B., Desmazes, C., Melot, T.,
Peter, M., Kovar, H., Joubert, I., de Jong, P., Rouleau, G., Aurias, A.
and Thomas, G.
Gene fusion with an ETS DNA-binding domain caused by chromosome
translocation in human tumours
Nature 359 (6391), 162-165 (1992)
MEDLINE 92396239
JOURNAL Nature 359 (6391), 162-165 (1992)
PUBMED 1522903
REFERENCE 2 (bases 1 to 2371)
Delattre, O.
Direct Submision
Submitted (26-MAY-1992) O. Delattre, Lab. de Genet. des Tumeurs.
Inst. Curie, 26 rue D'Ulm, 75231 Paris Cedex, FRANCE
On Sep 23, 1994 this sequence version replaced gi:1279.
COMMENT
FEATURES
source
1..2390
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/chromosome="2q12"
/clone_id="BPIA5"
/issue_type="brain"
/clone_lib="cDNA, Stratagene 936206"
/dev_stage="foetus"
1..2390
/gene="EMS"
44..2014
/gene="EMS"
/codon_start=1
/product="RNA binding protein"
/protein_id="CAA47350.1"
/db_xref="GI:31280"
/db_xref="GOA:Q01844"
/db_xref="UniProt/Swiss-Prot:Q01844"
/translation="MASTVSTYSQAAAGGSAVYAPOTGVAOTTOAYGQOISYGY
GQPTDVSYQAOATTAATGOTATATATGCGPTGCTTPTAQAAYGQVQGTGAVDTT
ATVTTQASVYAAQSAIGTQPAVAGQAPAAATPAPDQNKTEISQPSSTGGTQNO
PSLGYQSNYSYPOVPASYPQVPTAIPSPYPTSYSTQPTSYDSYSQOHTGQPS
SYQOQSYQOQSYGQOQPTSYPPQTSYQSOASQYQSOQSSSYQOQSSSFQDHPSSG
VYQOQSSYQOQSSYQOQSSYQOQSSYQOQSSYQOQSSYQOQSSYQOQSSYQOQSSY
KQSGPMDGPDLDLIGPVDPDSDNSALVYQGLNSVTLDLADFFKQGVYKXMR
TQGPPIHILIDKETGKPKDAYSIDPPTAKAAYVWFGKDKQSGKLYSLARKMR
MNSWRGLPPEBERGMPPLRGSPGSPGSPGSPGSPGSPGSPGSPGSPGSPGSPGSKRN
PSGGNVVORHAGDMQCPNPGCQNPANPAMTECQCAKPEGFLPPFPFGDGRG
GPGKMGKGRGLMDRGPGMGFRGGGGRGGGRGGGRGGGRGGGRGGGRGGGRGGP
LMEQMGRRGGRGGPGKMDKGEHRERRRBPY"
1127..1387
/gene="EMS"
/bound_moiety="RNA"
2162..2167
/gene="EMS"
/evidence=experimental
2350..2355
/gene="EMS"
/evidence=experimental
2371
/gene="EMS"
polya_site

```

## ORIGIN

Query Match 99.8%; Score 1010.4; DB 9; Length 2390; Best Local Similarity 99.9%; Pred. No. 1,4e-196; Matches 1011; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

1 AGGACGGGGTGGAAATGGGACCGCTGGAGCGAGGTGGCTTCAATTAAGCTGGTGGACC 60  
 DB 1000 AGGACGGGGTGGAAATGGGACCGCTGGAGCGAGGTGGCTTCAATTAAGCTGGTGGACC 1059

61 CATGATGAAGACACAGATCTTGATCTAGGCCCACTGTAGATCCAGATGAAGACTCTGA 120  
 DB 1060 CATGATGAAGACACAGATCTTGATCTAGGCCCTCTGTAGATCCAGATGAAGACTCTGA 1119

121 CAACAGTGCATTTATGTACAGAGATTAATGACAGTGTGACTCTAGATGATCTGGCAGA 180  
 DB 1120 CAACAGTGCATTTATGTACAGAGATTAATGACAGTGTGACTCTAGATGATCTGGCAGA 1179

181 CTTCTTAAAGCAGTGGGGTGTGTAGATGAACAAGAACTGGGGCAACCAATGATCCA 240  
 DB 1180 CTTCTTAAAGCAGTGGGGTGTGTAGATGAACAAGAACTGGGGCAACCAATGATCCA 1239

241 CATCTACCTGACACAGAGAAACAGAGAAAGCCCAAGCGATGCGCACTGTCTATGAAGA 300  
 DB 1240 CATCTACCTGACACAGAGAAACAGAGAAAGCCCAAGCGATGCGCACTGTCTATGAAGA 1299

301 CCCACCACTGCGCAAGGCTGCGTGGATGATGAGAAAGATTTTCAAGGAGCAAA 360  
 DB 1300 CCCACCACTGCGCAAGGCTGCGTGGATGATGAGAAAGATTTTCAAGGAGCAAA 1359

361 ACTTAAAGTCTCCCTGCTGGGAAAGAGCTCCAAATGAAGATGAGGGGTGCTGCC 420  
 DB 1360 ACTTAAAGTCTCCCTGCTGGGAAAGAGCTCCAAATGAAGATGAGGGGTGCTGCC 1419

421 ACCCGGTGAGGGGACAGAGGATGCGCACTCCGTGAGAGTCCAGAGAGCCCAAGAGG 480  
 DB 1420 ACCCGGTGAGGGGACAGAGGATGCGCACTCCGTGAGAGTCCAGAGAGCCCAAGAGG 1479

481 TCCGTGGGGACCCATGGGTTGCATGGAGGCGGTGAGAGATGAGAGGCTTCCCTCC 540  
 DB 1480 TCCGTGGGGACCCATGGGTTGCATGGAGGCGGTGAGAGATGAGAGGCTTCCCTCC 1539

541 AAGAGGACCCCGGGGTTCCCGAGGGGAAACCCCTCTGAGAGAGAGAACTCCAGACCGAGC 600  
 DB 1540 AAGAGGACCCCGGGGTTCCCGAGGGGAAACCCCTCTGAGAGAGAGAACTCCAGACCGAGC 1599

601 TGGAGACTGGCAGTGTCCCAATCCGGGTGTGAGAAACAGAACTTCCCTGAGAGACAGA 660  
 DB 1600 TGGAGACTGGCAGTGTCCCAATCCGGGTGTGAGAAACAGAACTTCCCTGAGAGACAGA 1659

661 GTGCAACCAAGTGAAGGCCCCCAAGCCTGAAAGCTTCTCCGCGCACCTTTTCGCCCC 720  
 DB 1660 GTGCAACCAAGTGAAGGCCCCCAAGCCTGAAAGCTTCTCCGCGCACCTTTTCGCCCC 1719

721 GGGTGGTGAATCGTGGAGAGAGTGGCCCTGTGTGAGTACGCGGGAGAGAAAGAGTGGCTCAT 780  
 DB 1720 GGGTGGTGAATCGTGGAGAGAGTGGCCCTGTGTGAGTACGCGGGAGAGAAAGAGTGGCTCAT 1779

781 GATCTGTGTGTCTCCCGTGAATGTTCAGAGGTGCGGTGTGTGAGACAGAGGTGGCTT 840  
 DB 1780 GATCTGTGTGTCTCCCGTGAATGTTCAGAGGTGCGGTGTGTGAGACAGAGGTGGCTT 1839

841 CCGTGTGCGCGGGGACATGAGCCGAGGTGCTTTGTGAGAGAAAGAGAGTGGCTTGG 900  
 DB 1840 CCGTGTGCGCGGGGACATGAGCCGAGGTGCTTTGTGAGAGAAAGAGAGTGGCTTGG 1899

901 GGGGCCCCCTTGGACCTTTGATGAGACAGATGGAGAGAAAGAGAGAGAGTGGAGACC 960  
 DB 1900 GGGGCCCCCTTGGACCTTTGATGAGACAGATGGAGAGAAAGAGAGAGAGTGGAGACC 1959

961 TGGAAAAATGATTAAGGCGAGCACCTCAGAGCGCAGAGATCGGCTTAC 1012  
 DB 1960 TGGAAAAATGATTAAGGCGAGCACCTCAGAGCGCAGAGATCGGCTTAC 2011

## ORIGIN

Query Match 99.7%; Score 1008.8; DB 6; Length 2026; Best Local Similarity 99.8%; Pred. No. 3.1e-196; Matches 1010; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

1 AGGACCGGTGGAAATGGGACCGCTGGAGCGAGGTGGCTTCAATTAAGCTGGTGGACC 60  
 DB 838 AGGACCGGTGGAAATGGGACCGCTGGAGCGAGGTGGCTTCAATTAAGCTGGTGGACC 897

61 CATGATGAAGACACAGATCTTGATCTAGGCCCACTGTAGATCCAGATGAAGACTCTGA 120  
 DB 898 CATGATGAAGACACAGATCTTGATCTAGGCCCACTGTAGATCCAGATGAAGACTCTGA 957

121 CAACAGTGCATTTATGTACAGAGATTAATGACAGTGTGACTCTAGATGATCTGGCAGA 180  
 DB 958 CAACAGTGCATTTATGTACAGAGATTAATGACAGTGTGACTCTAGATGATCTGGCAGA 1017

181 CTTCTTAAAGCAGTGGGGTGTGTAGATGAACAAGAACTGGGGCAACCATGATCCA 240  
 DB 1018 CTTCTTAAAGCAGTGGGGTGTGTAGATGAACAAGAACTGGGGCAACCATGATCCA 1077

241 CATCTACCTGACACAGAGAAACAGAGAAAGCCCAAGCGATGCGCACTGTCTATGAAGA 300  
 DB 1078 CATCTACCTGACACAGAGAAACAGAGAAAGCCCAAGCGATGCGCACTGTCTATGAAGA 1137

301 CCCACCACTGCGCAAGGCTGCGTGGATGATGAGAAAGATTTTCAAGGAGCAAA 360  
 DB 1138 CCCACCACTGCGCAAGGCTGCGTGGATGATGAGAAAGATTTTCAAGGAGCAAA 1197

361 ACTTAAAGTCTCCCTGCTGGGAAAGAGCTCCAAATGAAGATGAGGGGTGCTGCC 420  
 DB 1198 ACTTAAAGTCTCCCTGCTGGGAAAGAGCTCCAAATGAAGATGAGGGGTGCTGCC 1257

421 ACCCGGTGAGGGGACAGAGGATGCGCACTCCGTGAGAGTCCAGAGAGCCCAAGAGG 480  
 DB 1258 ACCCGGTGAGGGGACAGAGGATGCGCACTCCGTGAGAGTCCAGAGAGCCCAAGAGG 1317

481 TCCGTGGGGACCCATGGGTGCAATGGAGGCGGTGAGAGAGATTAAGAGAGCTTCCCTCC 540  
 DB 1318 TCCGTGGGGACCCATGGGTGCAATGGAGGCGGTGAGAGAGATTAAGAGAGCTTCCCTCC 1377

541 AAGAGGACCCCGGGGTTCCCGAGGGGAAACCTCTTGAAGAGAGAAAGTCCAGACCGAGC 600  
 DB 1378 AAGAGGACCCCGGGGTTCCCGAGGGGAAACCTCTTGAAGAGAGAAAGTCCAGACCGAGC 1437

601 TGGAGACTGGCAGTGTCCCAATCCGGGTGTGAGAAACAGAACTTCCCTGAGAGACAGA 660  
 DB 1438 TGGAGACTGGCAGTGTCCCAATCCGGGTGTGAGAAACAGAACTTCCCTGAGAGACAGA 1497

ORIGIN	Query Match	99.7%;	Score 1008.8;	DB 9;	Length 2026;
	Best Local Similarity	99.8%;	Pred. No. 3.1e-196;		
	Matches 1010;	Conservative	0;	Mismatches 2;	Indels 0;
				Gaps	0;
QY	1	AGNACGCGGTGGAATGCGCAGCGCTGGAGAGCGAGGTGCGTTCATTAAGCCTGTGTGACC	60		
Db	838	AGGACGCGGTGGAATGCGCAGCGCTGGAGAGCGAGGTGCGTTCATTAAGCCTGTGTGACC	897		
QY	61	CATGATGAAAGACACAGATCTTGATCTTAGGCCCACTGTAGTTCAGATGAAGCTCTGA	120		
Db	898	CATGATGAAAGACACAGATCTTGATCTTAGGCCCACTGTAGTTCAGATGAAGCTCTGA	957		
QY	121	CAACAGTGCAAATTATGTACAAAGATTAATATGACAGTGTACTCTAGATGTCGGCAGA	180		
Db	958	CAACAGTGCAAATTATGTATACAAAGATTAATATGACAGTGTACTCTAGATGTCGGCAGA	1017		
QY	181	CTTCTTTAAGCAGTGTGGGTGTGTTAATGATGAACAAAGAACTGGGCAACCCATGATCCA	240		
Db	1018	CTTCTTTAAGCAGTGTGGGTGTGTTAATGATGAACAAAGAACTGGGCAACCCATGATCCA	1077		
QY	241	CATTTACCTGGACAAAGAAACAGGAAAGCCCAAGGAGATGCCACAGTGTCTTATGAAGA	300		
Db	1078	CATTTACCTGGACAAAGGAAACAGGAAAGCCCAAGGAGATGCCACAGTGTCTTATGAAGA	1137		
QY	301	CCGACCCACATGCGCAAGGCTGCGGTGAATGATGATGGGAAAGATTTTCAAGGGAGCAA	360		
Db	1138	CCGACCCACATGCGCAAGGCTGCGGTGAATGATGATGGGAAAGATTTTCAAGGGAGCAA	1197		
QY	361	ACTTAAAGTCTCCCTTGCTCGGAAAGAGCTCCATGAACATATGCGGGGTGTCTTGCC	420		
Db	1198	ACTTAAAGTCTCCCTTGCTCGGAAAGAGCTCCATGAACATATGCGGGGTGTCTTGCC	1257		
QY	421	ACCCCGTAGAGGCGAGAGGCAATGCCACCACTCCGTGGAGGTCCAGAGGCCCAAGAGG	480		
Db	1258	ACCCCGTAGAGGCGAGAGGCAATGCCACCACTCCGTGGAGGTCCAGAGGCCCAAGAGG	1317		
QY	481	TCCGTGGGGGAGCCCATGTGGGTGCGCATGTGGAGGCGCGTGAAGAGATATGAAGAGCTTCCCTCC	540		
Db	1318	TCCGTGGGGGAGCCCATGTGGGTGCGCATGTGGAGGCGCGTGAAGAGATATGAAGAGCTTCCCTCC	1377		
QY	541	AAGAGAACCCCGGGGTTTCCGAGGGAAACCCCTCTGAGAGAGGAAAGCTCCAGCACCGAGC	600		
Db	1378	AAGAGAACCCCGGGGTTTCCGAGGGAAACCCCTCTGAGAGAGGAAAGCTCCAGCACCGAGC	1437		
QY	601	TGGAAGACTGCGCAGTGTCCCAATCCGGGTTGTGAAACCAAGACTTGTGCTGGAGAAAGAG	660		
Db	1438	TGGAAGACTGCGCAGTGTCCCAATCCGGGTTGTGAAACCAAGACTTGTGCTGGAGAAAGAG	1497		
QY	661	GTGCAACCAAGTATAGGCCCCCAAAGCCTGGAAGCTTCTCCCGCACAACCTTTCGCCCCC	720		
Db	1498	GTGCAACCAAGTATAGGCCCCCAAAGCCTGGAAGCTTCTCCCGCACAACCTTTCGCCCCC	1557		
QY	721	GAGTGTGATCTGTGCGCAGAGGTGCGCTGTGTGCAATCGGGAGAGAAAGGTGCGCTCAT	780		
Db	1558	GAGTGTGATCTGTGCGCAGAGGTGCGCTGTGTGCAATCGGGAGAGAAAGGTGCGCTCAT	1617		
QY	781	GGATCTGTGTGTGTCCCGGTGGAATGTTCAAGGTGTGCGGTGTGTGGAACAGAGGTGTGCTT	840		
Db	1618	GGATCTGTGTGTGTCCCGGTGGAATGTTCAAGGTGTGCGGTGTGTGGAACAGAGGTGTGCTT	1677		
QY	841	CCGTGTGTGCGCGGGGCGATGACCCGAGGTGTGCTTGTGTGAGGAAGAGAGGTGCGCTGTG	900		
Db	1678	CCGTGTGTGCGCGGGGCGATGACCCGAGGTGTGCTTGTGTGAGGAAGAGAGGTGCGCTGTG	1737		
QY	901	GGGGCCCCCTTGAGACTTTGATGGAACAGATGGAGGAAGAAAGAGAGAGCTGTGAGAGCC	960		
Db	1738	GGGGCCCCCTTGAGACTTTGATGGAACAGATGGAGGAAGAAAGAGAGAGCTGTGAGAGCC	1797		
QY	961	TGGAAGAAATGGAATTAAGCGAGCAACCGTCAAGAGCGGAGATGTGGCGCTTAC	1012		

Db 1798 TGGAAAAATGATTAAGCGGAGCAGCTGACAGAGCGAGAGATGCGCCCTAC 1849

RESULT 12  
LOCUS C0730352 2326 bp DNA linear PAT 03-FEB-2004  
DEFINITION Sequence 16286 from Patent WO02068579.  
ACCESSION C0730352  
VERSION C0730352.1 GI:42303963  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE  
AUTHORS Venter, C.J., Adams, M.C., Li, P.W. and Myers, E.W.  
TITLE Kites, such as nucleic acid arrays, comprising a majority of  
humanexons or transcripts, for detecting expression and other uses  
thereof  
JOURNAL Patent: WO 02068579-A 16286 06-SEP-2002,  
PE Corporation (NY) (US)  
FEATURES  
source 1. .2326  
/organism="Homo sapiens"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:9606"

ORIGIN

Query Match 98.9%; Score 1001; DB 6; Length 2326;  
Best Local Similarity 99.9%; Pred. No. 1.2e-194;  
Matches 1012; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1 AGGACCGCGGTGGAATGGGCGCGCTGGAAGCGAGGTGCTTCAATAGCTGGTGAGC 60  
DB 952 AGGACCGCGGTGGAATGGGCGCGCTGGAAGCGAGGTGCTTCAATAGCTGGTGAGC 1011  
QY 61 CATGATGGAAGGACCAAGATCTTGATGAGGCGCACTGATGATCCAGATGAAGCTCTGA 120  
DB 1012 CATGATGGAAGGACCAAGATCTTGATGAGGCGCACTGATGATCCAGATGAAGCTCTGA 1071  
QY 121 CAACAGTGCATTTATGTAACAAGATTAATGACAGTGTACTCTAGATGATCTGGCAGA 180  
DB 1072 CAACAGTGCATTTATGTAACAAGATTAATGACAGTGTACTCTAGATGATCTGGCAGA 1131  
QY 181 CTCTCTTTAAGCAGTGTGGGCTTTAATGATGAACAAGAGATCGGGCAACCATGATCCA 240  
DB 1132 CTCTCTTTAAGCAGTGTGGGCTTTAATGATGAACAAGAGATCGGGCAACCATGATCCA 1191  
QY 241 CATCTACCTGGAAGAAGGAAACAGAAAGCCCAAGGCGATGCCAGATGCTCTTATGAAGA 300  
DB 1192 CATCTACCTGGAAGAAGGAAACAGAAAGCCCAAGGCGATGCCAGATGCTCTTATGAAGA 1251  
QY 301 CCCACCCACTGCAAGAGGCTGCGTGAATGTTGATGGAAGAGATTTTCAAGGAGCA 360  
DB 1252 CCCACCCACTGCAAGAGGCTGCGTGAATGTTGATGGAAGAGATTTTCAAGGAGCA 1311  
QY 361 ACTTAAAGTCTCCCTTGTCTGGAAGAAGCTTCAATGAACAGTATGCGGGGTGTCTGCC 420  
DB 1312 ACTTAAAGTCTCCCTTGTCTGGAAGAAGCTTCAATGAACAGTATGCGGGGTGTCTGCC 1371  
QY 421 ACCCGGTGAGGAGGAGGAGATGACCAACATCTCGTGAAGTCCAGAGAGCCCAAGAG 480  
DB 1372 ACCCGGTGAGGAGGAGGAGATGACCAACATCTCGTGAAGTCCAGAGAGCCCAAGAG 1431  
QY 481 TCCTGGGGGACCCATGGGTGCAATGGAAGGCGGTGAGAGATGAGAGAGCTTCCCTCC 540  
DB 1432 TCCTGGGGGACCCATGGGTGCAATGGAAGGCGGTGAGAGATGAGAGAGCTTCCCTCC 1491  
QY 541 AAGAGAGACCCCGGGGTTCCCGAGAGAACTCTTGAAGAGAGAACTCTCCAGACCGGAG 600  
DB 1492 AAGAGAGACCCCGGGGTTCCCGAGAGAACTCTTGAAGAGAGAACTCTCCAGACCGGAG 1551  
QY 601 TGGAGACTGCACTGTCCCAATCCGGGTTGTGGAACCAAGAACTTGGCCCTGAGAAACAGA 660

Db 1552 TGGAGACTGCACTGTCCCAATCCGGGTTGTGGAACCAAGAACTTGGCCCTGAGAAACAGA 1611

QY 661 GTGCAACCACTGTAAAGGCCCAAGGCTTGAAGGCTTCTCCGCGACCCCTTTCGCCGCC 720  
DB 1612 GTGCAACCACTGTAAAGGCCCAAGGCTTGAAGGCTTCTCCGCGACCCCTTTCGCCGCC 1671  
QY 721 GGGTGTGATCTGTGCAAGAGTGGCCCTGTGTCATGCGGGGAGAGAGAGTGGCTTCAT 780  
DB 1672 GGGTGTGATCTGTGCAAGAGTGGCCCTGTGTCATGCGGGGAGAGAGAGTGGCTTCAT 1731  
QY 781 GGATCGTGTGTGTCGGCGTGAATGTTCAAGAGTGGCCGCTGTGAGAGAGAGAGTGGCTT 840  
DB 1732 GGATCGTGTGTGTCGGCGTGAATGTTCAAGAGTGGCCGCTGTGAGAGAGAGAGTGGCTT 1791  
QY 841 -CCGTGTGTGTCGGCGTGAATGTTCAAGAGTGGCCGCTGTGAGAGAGAGAGTGGCTT 899  
DB 1792 CCGTGTGTGTCGGCGGATGAGACGAGTGTGCTTGTGTGAGAGAGAGAGTGGCTT 1851  
QY 900 GGGGGGCCCCCTGACCTTTGATGAACAGATGGAGAGAGAGAGAGAGAGAGAGAGAG 959  
DB 1852 GGGGGGCCCCCTGACCTTTGATGAACAGATGGAGAGAGAGAGAGAGAGAGAGAGAG 1911  
QY 960 CTGAAAAAATGATTAAGGCGAGACCGCTGAGAGAGGCGAGATGCGCCCTAC 1012  
DB 1912 CTGAAAAAATGATTAAGGCGAGACCGCTGAGAGAGGCGAGATGCGCCCTAC 1964

RESULT 13  
LOCUS BC072442 2164 bp mRNA linear PRI 30-JUN-2004  
DEFINITION Homo sapiens Ewing sarcoma breakpoint region 1, transcript variant  
B072442  
ACCESSION BC072442  
VERSION BC072442.1 GI:48734726  
KEYWORDS  
SOURCE MGC.  
ORGANISM Homo sapiens (human)  
Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE  
AUTHORS Strausberg, R.L., Feingold, E.A., Grouse, L.H., Deng, J.G.,  
Klausner, R.D., Collins, P.S., Wagner, L., Shennan, C.M., Schuler, G.D.,  
Altschul, S.F., Zeeberg, B., Buecaw, K.H., Schaefer, C.F., Bhat, N.K.,  
Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Heish, F.,  
Diachenko, L., Marcovina, K., Farmer, A.A., Rubin, G.M., Hong, L.,  
Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L.,  
Schetz, T.E., Brownstein, M.J., Udell, T.B., Toshynski, S.,  
Carninci, P., Prange, C., Raha, S.S., Loggellano, N.A., Peters, G.J.,  
Abramson, R.D., Mullany, S.J., Bosak, S.A., McEwan, P.J.,  
McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S.,  
Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Ruhl, S.W.,  
Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,  
Fahy, J., Helton, E., Kettelman, M., Madan, A., Rodriguez, S.,  
Sanchez, A., Whitting, M., Madan, A., Young, A.C., Shevchenko, Y.,  
Boutard, G.G., Blakeley, R.W., Touchman, J.W., Green, E.D.,  
Dickson, M.C., Rodriguez, A.C., Grimmond, J., Schmitz, J., Myers, R.M.,  
Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Small, D.E.,  
Scherer, A., Schein, J.E., Jones, S.J. and Marra, M.A.  
TITLE Generation and initial analysis of more than 15,000 full-length  
human and mouse cDNA sequences  
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)  
PUBMED 12477932  
REFERENCE 2 (bases 1 to 2164)  
AUTHORS Strausberg, R.  
TITLE Direct Submision  
JOURNAL Submitted (01-JUN-2004) National Institutes of Health, Mammalian  
Gene Collection (MGC), Cancer Genomics Office, National Cancer  
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
USA  
REMARK NIH-MGC Project URL: <http://mgc.ncl.nih.gov>  
COMMENT Contact: MGC help desk  
Email: [cgabbs-remail.nih.gov](mailto:cgabbs-remail.nih.gov)



Oy 1 AGAAGCCGATGAAATGGGCAAGCGCTGAGAGAGGAGAGTGGCTTCATTAAGCTCGTGGAAAC 60  
 Db 960 AGAAGAGACCGGTGGAAATGGGCGCTGAGAGAGGAGAGTGGCTTCATTAAGCTCGTGGAAAC 1015  
 Oy 61 CATGATGAAGGACACGATCTTGATCTAGGCCCACTGTAGATCCAGATGAAGACTCTGA 120  
 Db 1020 CATGATGAAGGACACGATCTTGATCTAGGCCCACTGTAGATCCAGATGAAGACTCTGA 1075  
 Oy 121 CAACGTGCATTATTGTCACAGAGTTAAATGACGTGTGACTTGAATGATCTGGCAGA 180

Db	1080	CAACAGGCAATTATTAATGTCACAGAAATTAATATGACAGTGTACCTCTAGATGATCTGGGACA	1139
Qy	181	CTTCTTTAAGCAGTGGGGTTGTATAGATGAACAAGAGAACTGGGCAACCCATGATCCA	240
Db	1140	CTTCTTTAAGCAGTGGGGTTGTATAGATGAACAAGAGAACTGGGCAACCCATGATCCA	1139
Qy	241	CATCTACCTTGAGACAGAGAAACAGAAAGCCCAAGGCGATGCAAGTGTCTTAGAAGA	300
Db	1200	CATCTACCTTGAGACAGAGAAACAGAAAGCCCAAGGCGATGCAAGTGTCTTAGAAGA	1259
Qy	301	CCCAACCACTGCGCAAGGCTGGCGTGAATGGTTTGAATGGGAAAGATTTTCAAGGAGCAAA	360
Db	1260	CCCAACCACTGCGCAAGGCTGGCGTGAATGGTTTGAATGGGAAAGATTTTCAAGGAGCAAA	1319
Qy	361	ACTTAAAGTCTCCCTTGCTGCGAAGAGCCTCCAAATGAACGATATGCGGGGTGCTTGGCC	420
Db	1320	ACTTAAAGTCTCCCTTGCTGCGAAGAGCCTCCAAATGAACGATATGCGGGGTGCTTGGCC	1379
Qy	421	ACCCCGTAGGAGGACAGAGCATGCGACCACTCCGTGAGAGGTCCAGAGAGCCACAGAGG	480
Db	1380	ACCCCGTAGGAGGACAGAGCATGCGACCACTCCGTGAGAGGTCCAGAGAGCCACAGAGG	1439
Qy	481	TCCGTGGGGAGCCCATGGGTGTGCATATGGGAGGCGGTGAGAGAGATATAGAGAGGCTTCCCTCC	540
Db	1440	TCCGTGGGGAGCCCATGGGTGTGCATATGGGAGGCGGTGAGAGAGATATAGAGAGGCTTCCCTCC	1499
Qy	541	AAGAGGACCCCGGGGTTTCCCGAAGGAAACCCCTCTGAGAGAGAAACGTCCAGCACCGAGC	600
Db	1500	AAGAGGACCCCGGGGTTTCCCGAAGGAAACCCCTCTGAGAGAGAAACGTCCAGCACCGAGC	1559
Qy	601	TGAGAGACTGGCAGATGTCCCAATCCGGGTTGTGAAAACAGAACTTGGCCTTGAGAGACAGA	660
Db	1560	TGAGAGACTGGCAGATGTCCCAATCCGGGTTGTGAAAACAGAACTTGGCCTTGAGAGACAGA	1619
Qy	661	GTGCAACCAATGTATAGGCCCCCAAGCCTGGAAGCTTCTCCCGGCAACCTTTCCGCCCCC	720
Db	1620	GTGCAACCAATGTATAGGCCCCCAAGCCTGGAAGCTTCTCCCGGCAACCTTTCCGCCCCC	1679
Qy	721	GGGTGTGATCTGTGGCAGAGGTGGCCCTGTGTGCAATCCGGGAGGAAAGAGTGGCCTTCAT	780
Db	1680	GGGTGTGATCTGTGGCAGAGGTGGCCCTGTGTGCAATCCGGGAGGAAAGAGTGGCCTTCAT	1739
Qy	781	GGATCGAGTGTGTCCCGGTGGAATGTTCAGAGGTGGCCGTGTGTGAGACAGAGGTGACTT	840
Db	1740	GGATCGAGTGTGTCCCGGTGGAATGTTCAGAGGTGGCCGTGTGTGAGACAGAGGTGACTT	1799
Qy	841	CCGTGTGTGGCCCGGGGACATGAGCCGAGGTGGCTTTGTGTGAGAGAAACGAGGTGGCCCTTGG	900
Db	1800	CCGTGTGTGGCCCGGGGACATGAGCCGAGGTGGCTTTGTGTGAGAGAAACGAGGTGGCCCTTGG	1859
Qy	901	GGGGCCCCCTTGACCTTTGATGTGAACAGATGGGAGGAAAGAGGAGGAGAGGTGAGGAGACC	960
Db	1860	GGGGCCCCCTTGACCTTTGATGTGAACAGATGGGAGGAAAGAGGAGGAGAGGTGAGGAGACC	1919
Qy	961	TGGAATAATGATTAAGCGACGACCCCTCAGAGCGCAGAGATCGGCTTAC	1012
Db	1920	TGGAATAATGATTAAGCGACGACCCCTCAGAGCGCAGAGATCGGCTTAC	1971
RESULT 14			
AK056309	2189 bp	mRNA	linear
LOCUS	AK056309	Homo sapiens cDNA FLJ11747 fis, clone NT2R12007377, highly similar	PRI 30-JAN-2004
DEFINITION	to RNA-BINDING PROTEIN EWS.		
ACCESSION	AK056309		
VERSION	AK056309.1	GI:16551673	
KEYWORDS	c1igo capping; fis (full insert sequence).		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryotes; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
AUTHORS	Oca,T., Suzuki,Y., Nishikawa,T., Oseuki,T., Sugiyama,T., Irie,R.,		



Makamatsu, A., Hayashi, K., Sato, H., Nagai, K., Kimura, K., Matita, H., Sakita, M., Obayashi, M., Mishi, T., Shibahara, T., Tanaka, T., Ishii, S., Yamamoto, J., Saito, K., Kawai, Y., Isono, Y., Nakamura, Y., Nagahara, K., Murakami, K., Yasuda, T., Iwayanagi, T., Watanabe, M., Shiratori, A., Sudo, H., Hosoi, T., Kaku, Y., Kodaira, H., Kondo, H., Sugawara, M., Takahashi, M., Kanda, K., Yokoi, T., Furuya, T., Kikkawa, E., Omura, Y., Abe, K., Kamihara, K., Katsura, N., Sato, K., Tanikawa, M., Yamazaki, M., Nimomiya, K., Ishibashi, T., Yamashita, H., Murakawa, K., Fujimori, K., Tanai, H., Kimata, M., Watanabe, M., Hirooka, S., Chiba, Y., Ishida, S., Ono, Y., Takiguchi, S., Watanabe, S., Yoshida, M., Hotta, T., Kusano, J., Kanehori, K., Takahashi, F., Hara, H., Tanase, T., Nomura, Y., Togita, S., Komai, F., Hara, R., Takeuchi, K., Arita, M., Imose, N., Mutsaers, K., Yuki, H., Ohima, A., Sasaki, N., Aotsuka, S., Yoshikawa, Y., Matsunawa, H., Ichihara, T., Shiohara, N., Sano, S., Moriya, S., Komiyama, H., Satoh, N., Takami, S., Terashima, Y., Suzuki, O., Nakagawa, S., Senoh, A., Mizoguchi, H., Goto, Y., Shimizu, F., Wakebe, H., Hishigaki, H., Watanabe, T., Sugiyama, A., Takemoto, M., Kawakami, B., Yamazaki, M., Watanabe, K., Kumagai, A., Itakura, S., Fukuzumi, Y., Fujimori, Y., Komiyama, M., Tashiro, H., Tanigami, A., Fujiwara, T., Ono, T., Yanada, K., Fujii, Y., Ozaki, K., Hiro, M., Ohmori, Y., Kawabata, A., Hiki, T., Kobatake, N., Inagaki, H., Ikema, Y., Okamoto, S., Okitani, R., Kawakami, T., Noguchi, S., Itoh, T., Shigeta, K., Senba, T., Matsumura, K., Nakajima, Y., Mizuno, T., Morinaga, M., Sasaki, M., Togaishi, T., Oyama, M., Hata, H., Watanabe, M., Komatsu, T., Mizushima-Sugano, J., Sato, T., Shirai, Y., Takahashi, Y., Nakagawa, K., Okumura, K., Nagase, T., Nomura, N., Kikuchi, H., Masuno, Y., Yamashita, R., Nakai, K., Yada, T., Nakamura, Y., Ohara, O., Isogai, T. and Sugano, S. Complete sequencing and characterization of 21,243 full-length human cDNAs Nat. Genet. 36 (1), 40-45 (2004) 14702039

**TITLE**  
**JOURNAL**  
**PUBLISHED**  
**REFERENCE**  
**AUTHORS**

2  
1shibashi, T., Kanehori, K., Yoshida, M., Watanabe, S., Ishida, S., Ono, Y., Hotta, T., Hirooka, S., Murakami, K., Takiguchi, S., Kusano, J., Watanabe, M., Fujimori, K., Tanai, H., Ishida, M., Yamashita, H., Chiba, Y., Sugiyama, T., Irie, R., Ozaki, T., Sato, H., Sakita, M., Mishi, K., Yamamoto, J., Isono, Y., Kawai, H., Sato, H., Sekine, M., Kikuchi, H., Kanda, K., Matsumura, M., Takahashi, F., Ohima, A., Sugiyama, A., Kawakami, B., Suzuki, Y., Sugano, S., Nagahara, K., Masuno, Y., Nagai, K. and Isogai, T. NEDO human cDNA sequencing project Unpublished

**TITLE**  
**JOURNAL**  
**REFERENCE**  
**AUTHORS**  
**TITLE**  
**JOURNAL**

3 (bases 1 to 2189)  
Isogai, T., Otsuki, T. and Sugiyama, T. Direct Submission Submitted (24-OCT-2001) Takao Isogai, Helix Research Institute, Genomics Laboratory, 1532-3 Yano, Kisarazu, Chiba 252-0812, Japan (E-mail: isogai@helix.co.jp, Tel: 81-438-52-3975, Fax: 81-438-52-3986) NEDO human cDNA sequencing project supported by Ministry of Economy, Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology (RAB); cDNA library construction: Helix Research Institute (HRI) (supported by Japan Key Technology Center etc.); 5'-3' end one pass sequencing: RAB, HRI, and Biotechnology Center, National Institute of Technology and Evaluation; clone selection for full insert sequencing: RAB and HRI.

**FEATURES**  
**source**

1. 2189  
Location/Qualifiers  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="NT2R12007377"  
/cell\_type="NT2"  
/cell\_line="keratinocyte"  
/note="cloning vector: pMT189FL3-mRNA from NT2 neuronal precursor cells treated 2-weeks mitotic inhibitor after 5-weeks retinoic acid (RA) induction.-majority NT2 neuron"  
22. 2007  
/note="unnamed protein product"  
/codon\_start=1

/protein\_id="BA871145.1"  
/db\_xref="GI:16551674"  
/translation="MASTDYSTYSSQAAGQGSAYTAQPTGVAQTAQYAGQGSYGT  
GQPTDVSYPDQATTAATGQVAVTSTGQPTVGTSTGTATPAQVAPVQSYGTG  
AYPTTATVTTTQASVAAAGSAYGTGPAVPGQGPATATPRTDQKPTSTPOSS  
TGYVNPSTGSGSAYTPVPGSYTQAPATPAPSPSTSTPTSTPTSTPTSTPTST  
TYQDPSYQGSSTYQGSSTYQGPPTSPPTQTSYQAPBQYQGSSTYQGSSTYQGS  
HESKATGQESGSGSGPGENRMSGPDNDKRGSGDRGMSRGSGSGSGSGSGSG  
RGFNKRGSGPMDGPDLDLPPVDPEDSDNSAIYQGLNDVTLDDLDFQCGV  
KMKRTGQPMWTHLYLDEKGTGPKGDAVSEYDPTKAAVEMFDGDFQSKLVSLA  
RKRPNNMSRGGLPPEKRGMPPLKRGPGPGPGPGPGPGPGPGPGPGPGPGPG  
GSGNPGSG  
DRRGSGPGSG  
GPGPLMEQMGSGRGRGSGGKMDKSHHQRDRRY"

## ORIGIN

Query Match 98.4%; Score 996; DB 9; Length 2189;  
Best Local Similarity 98.8%; Pred. No 1,3e-193;  
Matches 1000; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

1 AGGACGGCGTGAATGCGGACGCGCTGAGAGCGAGTGCCTCAATAGCCTGTGAC 60  
993 AGGAGACGGCGTGAATGCGGACGCGCTGAGAGCGAGTGCCTCAATAGCCTGTGAC 1052  
61 CATGATGAAAGACCGAGATTTGATCTAGGCGCACTGTGATCCAGATGAACCTCTGA 120  
1053 CATGATGAAAGACCGAGATTTGATCTAGGCGCACTGTGATCCAGATGAACCTCTGA 1112  
121 CAACAGGCAATTTATGTAACAAGATTTAATGACAGTGAAGCTGTGATGATCTGCGAG 180  
1113 CAACAGGCAATTTATGTAACAAGATTTAATGACAGTGAAGCTGTGATGATCTGCGAG 1172  
181 CTTCTTAAGCAGTGTGGGTTGTTAAGATGAACAAGAACTGGGCAACCCATGATCCA 240  
1173 CTTCTTAAGCAGTGTGGGTTGTTAAGATGAACAAGAACTGGGCAACCCATGATCCA 1232  
241 CATCTACCTGAGCAAGAAACAGAAAGCCCAAGGCGATGCCACAGTGTCTATGAGA 300  
1233 CATCTACCTGAGCAAGAAACAGAAAGCCCAAGGCGATGCCACAGTGTCTATGAGA 1292  
301 CCCACCACTGCGCAAGGCTGCGTGGAAATGTTGATGGGAAAGATTTCAAGGAGCA 360  
1293 CCCACCACTGCGCAAGGCTGCGTGGAAATGTTGATGGGAAAGATTTCAAGGAGCA 1352  
361 ACTTAAGTCTCCCTTCTCGAAGAAAGCTCCATGAACAGTATCGGGTGTCTGCC 420  
1353 ACTTAAGTCTCCCTTCTCGAAGAAAGCTCCATGAACAGTATCGGGTGTCTGCC 1412  
421 ACCCGTGAAGGAGGAGGATGCGACCACTCGGTGAGTTCAGAGAGCCCAAGAG 480  
1413 ACCCGTGAAGGAGGAGGATGCGACCACTCGGTGAGTTCAGAGAGCCCAAGAG 1472  
481 TCCGGGGGACCAATGGTTCGATGAGGAGGCGTGGAGAGATAGAGAGGCTTCCCTCC 540  
1473 TCCGGGGGACCAATGGTTCGATGAGGAGGCGTGGAGAGATAGAGAGGCTTCCCTCC 1532  
541 AAGAGGACCCCGGGGTTCCCGAGGAGAACCTCTGAGAGAGAGAGTCCAGACCGAG 600  
1533 AAGAGGACCCCGGGGTTCCCGAGGAGAACCTCTGAGAGAGAGAGTCCAGACCGAG 1592  
601 TGGAGACTGGCAGTGTCCCAATCCGGGTTGTGAGAACAGAACTTGCCTGAGAACAG 660  
1593 TGGAGACTGGCAGTGTCCCAATCCGGGTTGTGAGAACAGAACTTGCCTGAGAACAG 1652  
661 GGGCAACGATGTAAGGCGCCCAAGGCTTCCCGGCACTCTTCCGCGCC 720  
1653 GGGCAACGATGTAAGGCGCCCAAGGCTTCCCGGCACTCTTCCGCGCC 1712  
721 GGGTGTGATGTCGTCAGAGTGTGCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 780  
1713 GGGTGTGATGTCGTCAGAGTGTGCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 1772  
781 GGATCGTGTGTCCTCCGCTGAGATGTTTCAAGAGTGGCGTGTGAGACAGAGTGTGCT 840



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using bw model

Run on: February 20, 2005, 00:31:32 / Search time 490.802 Seconds  
(without alignment)  
12206.093 Million cell updates/sec

Title: US-10-791-017a-1\_COPY\_1000\_2011

Perfect score: 1012

Sequence: 1 aggcagcggtgagatgsgca.....agcgagagatcgccctac 1012

Scoring table: IDENTITY\_NUC

Gapop 10.0, Gapext 1.0

Searched: 4390206 seqs, 295987067 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N\_Geneseq\_16Dec04:\*  
1: geneseqn1980s:\*  
2: geneseqn1990s:\*  
3: geneseqn2000s:\*  
4: geneseqn2001a:\*  
5: geneseqn2001b:\*  
6: geneseqn2002a:\*  
7: geneseqn2002b:\*  
8: geneseqn2003a:\*  
9: geneseqn2003b:\*  
10: geneseqn2003c:\*  
11: geneseqn2003d:\*  
12: geneseqn2004a:\*  
13: geneseqn2004b:\*

Pred. No. is the number of results predicted by chance a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1012	100.0	1988	10	ADA53506 Human cod
2	1012	100.0	2176	6	AA562262
3	1012	100.0	2372	13	AD086032 Human tum
4	1012	100.0	2390	13	AD816277 Human cdn
5	1010.4	99.8	2371	2	AAQ50643 Human Ews
6	1010.4	99.8	2388	5	AA570647 DNA encod
7	1010.4	99.8	2390	6	ABK84628 Human cdn
8	1010.4	99.8	2390	6	ABK64822 Human ben
9	1010.4	99.8	2390	6	ABN97274 Gene #377
10	1010.4	99.8	2390	13	ACN40903 Tumour-as
11	1008.8	99.7	2026	13	AD817446 Full leng
12	954.4	94.3	1807	13	ADP56333 Human PRO
13	908	89.7	2177	13	ACN37537 Tumour-as
14	882.4	87.2	2273	6	AA562623 cDNA sequ
15	879.2	86.9	2188	6	AB199383 Mouse ibc
16	653.6	56.3	1462	4	AA126725 Human bre
17	570	53.2	2299	11	ACN88697 Breast ca
18	538	50.2	3309	2	AAQ50646 Human Ews
19	508.4	43.2	568	9	ACH40478 Human foe
20	437.2			9	ACH22345 Human adu

## ALIGNMENTS

RESULT 1	AD53506	standard; cDNA, 1988 bp.
AD53506	AD53506	
AC	XX	
XX	XX	
DT	20-NOV-2003	(first entry)
XX	XX	
DE	Human coding sequence, SEQ ID 1074.	
XX	XX	
XX	Cytostatic; Anti-inflammatory; Osteopathic; Neuroprotective; Nootropic;	
KM	Gene Therapy; human; secretory protein; membrane proteins; cancer;	
KM	Inflammatory disease; osteoporosis; neurological disease; gene; ss.	
OS	Homo sapiens.	
XX	XX	
PN	EP1293569-A2.	
XX	XX	
PD	19-MAR-2003.	
XX	XX	
PF	21-MAR-2002; 2002EP-00006586.	
XX	XX	
PR	14-SEP-2001; 2001JP-00328381.	
XX	XX	
PR	24-JAN-2002; 2002US-0350435P.	
PA	(HELI-) HELIX RES INST.	
PA	(REAS-) RES ASSOC BIOTECHNOLOGY.	
XX	XX	
PI	Iogai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S;	
PI	Yamanoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Tamehika I;	
PI	Seki N, Yoshikawa T, Otsuka M, Nagahara K, Masuno Y;	
XX	XX	
DR	WPI; 2003-395539/38.	
XX	XX	
XX	P-PSDB; ADA55145.	
PT	New polynucleotides encoding full-length polypeptides, e.g. secretory	
PT	and/or membrane proteins, useful for developing medicines for diseases in	
PT	which the gene is involved, or as target molecules for gene therapy.	
XX	XX	
PS	Claim 1; SEQ ID NO 1074; 205bp; English.	
XX	XX	
CC	The present invention relates to novel human secretory or membrane	
CC	proteins (ADA54072-ADA55710) and their coding sequences (ADA5433-	

21	419.2	41.4	471	2	AAV6051	AAV6051 EST clone
22	309	30.5	545	3	AACT8275	AACT8275 Human can
23	304.6	30.1	665	13	AD05454	AD05454 Novel can
24	269.8	26.7	361	8	ABX38686	ABX38686 Bovine ES
25	258.4	25.5	503	4	AA118034	AA118034 Probe #79
26	258.4	25.5	503	4	ABA63001	ABA63001 Human foe
27	258.4	25.5	503	4	AA143027	AA143027 Probe #11
28	258.4	25.5	503	4	ABA30262	ABA30262 Probe #87
29	258.4	25.5	503	4	AAK37201	AAK37201 Human bon
30	258.4	25.5	503	4	AAK11415	AAK11415 Human bra
31	258.4	25.5	503	4	AB536871	AB536871 Human liv
32	258.4	25.5	503	6	AB511189	AB511189 Human gen
33	258.4	25.5	503	12	ACH71268	ACH71268 Human gen
34	254.2	25.1	262	12	ACH84968	ACH84968 Human gen
35	253	25.0	253	4	AA127213	AA127213 Probe #17
36	253	25.0	253	4	ABA75494	ABA75494 Human foe
37	253	25.0	253	4	AA156067	AA156067 Probe #24
38	253	25.0	253	4	ABA40118	ABA40118 Probe #18
39	253	25.0	253	4	AAK50129	AAK50129 Human bon
40	253	25.0	253	4	AAK24064	AAK24064 Human bra
41	253	25.0	253	4	AB549761	AB549761 Human liv
42	253	25.0	253	6	AB523605	AB523605 Human gen
43	212.4	21.0	457	4	AAK53923	AAK53923 Murine tr
44	176.8	17.5	1832	12	AD126112	AD126112 Human cdn
45	176.6	17.5	684	13	ACN37539	ACN37539 Tumour-as

AD54072). The coding sequences are useful in the gene therapy of diseases caused by abnormalities of the proteins, e.g. cancer, inflammatory diseases, osteoporosis or neurological disease.

Sequence 1988 BP; 519 A; 504 C; 584 G; 381 T; 0 U; 0 Other;

Query Match 100.0%; Score 1012; DB 10; Length 1988;  
Best Local Similarity 100.0%; Pred. No. 2.9e-233; Indels 0; Gaps 0;  
Matches 1012; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 AGGACGGCGTGAATGGGACCGCTGGAGCGAGGTGGCTTCAATAGCTGGTGGACC 60
DB AGGACGGCGTGAATGGGACCGCTGGAGCGAGGTGGCTTCAATAGCTGGTGGACC 859
QY 61 CATGATGAAGAGCAGATCTTGATCTAGGCCCACTGTGATCCAGATGAAGAATCTGA 120
DB CATGATGAAGAGCAGATCTTGATCTAGGCCCACTGTGATCCAGATGAAGAATCTGA 919
QY 121 CAACAGTGCATTTATTAATGAAGATTAAATGACAGTGAATCTAGATGATGCGGAGA 180
DB CAACAGTGCATTTATTAATGAAGATTAAATGACAGTGAATCTAGATGATGCGGAGA 979
QY 181 CTCTTTAAGCAGTGTGGGTTGTTAAGATGAACAAGAACTGGGCAACCAATGATCA 240
DB CTCTTTAAGCAGTGTGGGTTGTTAAGATGAACAAGAACTGGGCAACCAATGATCA 1039
QY 241 CATCTACCTGGAACAAGAAACAGAAAGCCCAAGGATGCCAAGTGTCTTATGAAGA 300
DB CATCTACCTGGAACAAGAAACAGAAAGCCCAAGGATGCCAAGTGTCTTATGAAGA 1099
QY 301 CCCACCCACTGCGCAGGCTGCGTGAATGTTGATGGGAAAGATTTTAAAGGAGCA 360
DB CCCACCCACTGCGCAGGCTGCGTGAATGTTGATGGGAAAGATTTTAAAGGAGCA 1159
QY 1100 CCAACCACTGCGCAGGCTGCGTGAATGTTGATGGGAAAGATTTTAAAGGAGCA 1159
DB CCAACCACTGCGCAGGCTGCGTGAATGTTGATGGGAAAGATTTTAAAGGAGCA 1159
QY 361 ACTTAAAGTCTCCCTTCTCGAAGAACCTCCATGAACAGTATGCGGGTGTCTGCC 420
DB ACTTAAAGTCTCCCTTCTCGAAGAACCTCCATGAACAGTATGCGGGTGTCTGCC 1219
QY 421 ACCCGGTGAGGGCAGAGCATGCCACCACTCTGTGAGAGTCCAGAGGCCCAAGAG 480
DB ACCCGGTGAGGGCAGAGCATGCCACCACTCTGTGAGAGTCCAGAGGCCCAAGAG 1220
QY 1220 ACCCGGTGAGGGCAGAGCATGCCACCACTCTGTGAGAGTCCAGAGGCCCAAGAG 1279
DB ACCCGGTGAGGGCAGAGCATGCCACCACTCTGTGAGAGTCCAGAGGCCCAAGAG 1279
QY 481 TCCGTGGGGGACCCATGGGATGGGATGGAGGCGGTGAAGAGATGAAGAGGCTTCCCTCC 540
DB TCCGTGGGGGACCCATGGGATGGGATGGAGGCGGTGAAGAGATGAAGAGGCTTCCCTCC 1280
QY 1280 TCCGTGGGGGACCCATGGGATGGGATGGAGGCGGTGAAGAGATGAAGAGGCTTCCCTCC 1339
DB TCCGTGGGGGACCCATGGGATGGGATGGAGGCGGTGAAGAGATGAAGAGGCTTCCCTCC 1339
QY 541 AAGAGGACCCCGGGGTTCCCGAGGGAACCCCTCTGAGAGAGAAACGTCCAGACCGAGC 600
DB AAGAGGACCCCGGGGTTCCCGAGGGAACCCCTCTGAGAGAGAAACGTCCAGACCGAGC 1340
QY 1340 AAGAGGACCCCGGGGTTCCCGAGGGAACCCCTCTGAGAGAGAAACGTCCAGACCGAGC 1399
DB AAGAGGACCCCGGGGTTCCCGAGGGAACCCCTCTGAGAGAGAAACGTCCAGACCGAGC 1399
QY 601 TGAAGACTGGCAGTGTCCCAATCCGGGTTGTGAACCAAGAACTTGGCTGGAGAACGA 660
DB TGAAGACTGGCAGTGTCCCAATCCGGGTTGTGAACCAAGAACTTGGCTGGAGAACGA 1400
QY 1400 TGAAGACTGGCAGTGTCCCAATCCGGGTTGTGAACCAAGAACTTGGCTGGAGAACGA 1459
DB TGAAGACTGGCAGTGTCCCAATCCGGGTTGTGAACCAAGAACTTGGCTGGAGAACGA 1459
QY 661 GTGACCAAGTGAAGGCCCCAAGACCTGAAGGCTTCCCGCACCTTTCCGCCCC 720
DB GTGACCAAGTGAAGGCCCCAAGACCTGAAGGCTTCCCGCACCTTTCCGCCCC 1460
QY 1460 GTGACCAAGTGAAGGCCCCAAGACCTGAAGGCTTCCCGCACCTTTCCGCCCC 1519
DB GTGACCAAGTGAAGGCCCCAAGACCTGAAGGCTTCCCGCACCTTTCCGCCCC 1519
QY 721 GGGTGTGATCTGTGGCAGAGGTGGCCCTGTGTGCATCGGAGAGAAAGGTGGCTCAT 780
DB GGGTGTGATCTGTGGCAGAGGTGGCCCTGTGTGCATCGGAGAGAAAGGTGGCTCAT 1520
QY 1520 GGGTGTGATCTGTGGCAGAGGTGGCCCTGTGTGCATCGGAGAGAAAGGTGGCTCAT 1579
DB GGGTGTGATCTGTGGCAGAGGTGGCCCTGTGTGCATCGGAGAGAAAGGTGGCTCAT 1579
QY 781 GAATGTGTGTGTCCCGGTGAATGTTCAAGGTGGCCGTGTGTGAAGACAGAGTGGCTT 840
DB GAATGTGTGTGTCCCGGTGAATGTTCAAGGTGGCCGTGTGTGAAGACAGAGTGGCTT 1580
QY 1580 GAATGTGTGTGTCCCGGTGAATGTTCAAGGTGGCCGTGTGTGAAGACAGAGTGGCTT 1639
DB GAATGTGTGTGTCCCGGTGAATGTTCAAGGTGGCCGTGTGTGAAGACAGAGTGGCTT 1639
QY 841 CCGTGTGTGTGTCCCGGTGAATGTTCAAGGTGGCCGTGTGTGAAGACAGAGTGGCTT 900
DB CCGTGTGTGTGTCCCGGTGAATGTTCAAGGTGGCCGTGTGTGAAGACAGAGTGGCTT 1640
QY 1640 CCGTGTGTGTGTCCCGGTGAATGTTCAAGGTGGCCGTGTGTGAAGACAGAGTGGCTT 1659
DB CCGTGTGTGTGTCCCGGTGAATGTTCAAGGTGGCCGTGTGTGAAGACAGAGTGGCTT 1659
QY 901 GGGGCCCCCTTGACCTTTGATGAACAGATGGAGAGAAAGAGAGACTGAGAGACC 960
DB GGGGCCCCCTTGACCTTTGATGAACAGATGGAGAGAAAGAGAGACTGAGAGACC 1700
QY 1700 GGGGCCCCCTTGACCTTTGATGAACAGATGGAGAGAAAGAGAGACTGAGAGACC 1759
DB GGGGCCCCCTTGACCTTTGATGAACAGATGGAGAGAAAGAGAGACTGAGAGACC 1759

```

QY 961 TGGAAATATGATTAAGGCGAGCACCGCTCAGAGGCGCAGAGATCGGCCCTAC 1012  
DB TGGAAATATGATTAAGGCGAGCACCGCTCAGAGGCGCAGAGATCGGCCCTAC 1811

RESULT 2  
AAS62262/C  
ID AAS62262 standard; cDNA; 2176 BP.  
XX  
XX AAS62262;  
XX  
XX 14-FEB-2002 (first entry)  
XX

DE cDNA sequence #49 encoding novel human secreted protein.

XX Human secreted protein; hyperproliferative disorder; autoimmune disorder;  
XX immune deficiency disorder; blood disorder; inflammatory disorder;  
XX infectious disorder; gene therapy; antimicrobial; hepatotropic;  
XX immunosuppressive; antineumatic; ss.  
XX  
XX Homo sapiens.  
XX

OS  
XX  
XX MO200177291-A2.  
XX

PD 18-OCT-2001.

XX 29-MAR-2001; 2001MO-US010485.  
XX

PR 06-APR-2000; 2000US-0195604P.  
XX

XX (GENY ) GENETICS INST INC.  
XX

XX Wong GG, Clark HF, Fechtel K, Agostino MJ, Howes SH, Resnick RJ;  
XX Galukota K, Graham JR;  
XX

XX WPI; 2002-010900/01.  
XX

PT New polynucleotides encoding secreted proteins useful for treating e.g.  
PT asthma, HIV and Crohn's disease.  
XX

PS Claim 1; Page 100; 391pp; English.  
XX

XX The present invention relates to the isolation of novel cDNA sequences  
XX which encode human secreted proteins. The cDNA sequences have been  
XX derived from a variety of human tissues. The invention also provides a  
XX method for producing proteins from these polynucleotide sequences. The  
XX proteins are useful for identifying compounds that modulate their  
XX activity and production, and the cell is also useful for identifying  
XX compounds that modulate expression of the polynucleotide sequences  
XX encoding the secreted proteins. The sequences of the invention are useful  
XX for treating diseases such as hyperproliferative disorders (e.g. cancer),  
XX immune deficiency disorders (e.g. severe combined immunodeficiency  
XX (SCID)), autoimmune disorders (e.g. multiple sclerosis), blood disorders  
XX (e.g. thrombocytopenia), inflammatory disorders (e.g. arthritis) and  
XX infectious disorders (e.g. hepatitis). The polynucleotide sequences of  
XX the invention are also useful in gene therapy. AAS62214-AAS62038  
XX represent the cDNA sequences of the invention that encode for novel human  
XX secreted proteins  
XX

SO Sequence 2176 BP; 413 A; 617 C; 563 G; 583 T; 0 U; 0 Other;

Query Match 100.0%; Score 1012; DB 6; Length 2176;  
Best Local Similarity 100.0%; Pred. No. 3e-233;  
Matches 1012; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 AGGACGGCGTGAATGGGACCGCTGGAGCGAGGTGGCTTCAATAGCTGGTGGACC 60
DB AGGACGGCGTGAATGGGACCGCTGGAGCGAGGTGGCTTCAATAGCTGGTGGACC 1205
QY 1205 AGGACGGCGTGAATGGGACCGCTGGAGCGAGGTGGCTTCAATAGCTGGTGGACC 1146
DB 61 CATGATGAAGAGCAGATCTTGATCTAGGCCCACTGTGATCCAGATGAAGAATCTGA 120
DB CATGATGAAGAGCAGATCTTGATCTAGGCCCACTGTGATCCAGATGAAGAATCTGA 1086

```

QY 121 CAAAGTGAATTTATTAACAAGATTAAATGAAGTGTGACTTATGATGATCTGGAGA 180  
DB 1085 CAAAGTGAATTTATTAACAAGATTAAATGAAGTGTGACTTATGATGATCTGGAGA 1026  
QY 181 CTTCTTTAAGCAGTGTGGGTTGTAAAGATGAACAAGAGACTGGGCAACCCATGATCCA 240  
DB 1025 CTTCTTTAAGCAGTGTGGGTTGTAAAGATGAACAAGAGACTGGGCAACCCATGATCCA 966  
QY 241 CATCTAAGTGAACAAGAAACAGAAAGCCCAAGGCGATGCAAGTGTCTTATGAGA 300  
DB 965 CATCTAAGTGAACAAGAAACAGAAAGCCCAAGGCGATGCAAGTGTCTTATGAGA 906  
QY 301 CCCACCACTGCAAGGCTGCGGTGAATGTGTTGATGGGAAAGATTTTAAAGGAGCAA 360  
DB 905 CCCACCACTGCAAGGCTGCGGTGAATGTGTTGATGGGAAAGATTTTAAAGGAGCAA 846  
QY 361 ACTTAAAGTCTCCCTTGTGTAAGAAAGCTCCATGAACAGTATGCGGGTGGTCTGCC 420  
DB 845 ACTTAAAGTCTCCCTTGTGTAAGAAAGCTCCATGAACAGTATGCGGGTGGTCTGCC 786  
QY 421 ACCCGTGAAGGAGAGAGGATGCAACCACTCCGTGAAGTCCAGAGGCGCCAGAGAG 480  
DB 785 ACCCGTGAAGGAGAGAGGATGCAACCACTCCGTGAAGTCCAGAGGCGCCAGAGAG 726  
QY 481 TCTTGGGGGACCAATGGGTGCAATGAGAGGCGCTGAGAGAGATGAGAGAGCTTCTCC 540  
DB 725 TCTTGGGGGACCAATGGGTGCAATGAGAGGCGCTGAGAGAGATGAGAGAGCTTCTCC 666  
QY 541 AAGAGAGACCCGGGGTTCCCGAGAGGAAACCCCTCTGAGAGAGGAAACGTCAGACAGC 600  
DB 665 AAGAGAGACCCGGGGTTCCCGAGAGGAAACCCCTCTGAGAGAGGAAACGTCAGACAGC 606  
QY 601 TGGAGACTGCAAGTGTCCCAATCGGGTTGTGAAACAGAACTTCCCTGAGAGACAGA 660  
DB 605 TGGAGACTGCAAGTGTCCCAATCGGGTTGTGAAACAGAACTTCCCTGAGAGACAGA 546  
QY 661 GTGCAACCAAGTGTGAAGGCGCCCAAGGCTTCTCCCGCACCTTTCGCGCC 720  
DB 545 GTGCAACCAAGTGTGAAGGCGCCCAAGGCTTCTCCCGCACCTTTCGCGCC 486  
QY 721 GGGTGTGATCGTGGAGAGAGTGGCCCTGTGGATGCGGGGAGAAAGTGGCCCTCAT 780  
DB 485 GGGTGTGATCGTGGAGAGAGTGGCCCTGTGGATGCGGGGAGAAAGTGGCCCTCAT 426  
QY 781 GGAATGATGATGATCGCGGTGAATGTTCAAGAGTGGCCCTGTGAGAGACAGAGTGGCTT 840  
DB 425 GGAATGATGATGATCGCGGTGAATGTTCAAGAGTGGCCCTGTGAGAGACAGAGTGGCTT 366  
QY 841 CCGTGTGATCGCGGGGCAATGAGCGAGTGGCTTGTGAGAGAAACAGAGTGGCCCTTGG 900  
DB 365 CCGTGTGATCGCGGGGCAATGAGCGAGTGGCTTGTGAGAGAAACAGAGTGGCCCTTGG 306  
QY 901 GGGGCCCCCTTGAAGCTTGAAGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 960  
DB 305 GGGGCCCCCTTGAAGCTTGAAGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 246  
QY 961 TGGAAATGATGAAGGCGAGACCGTCAAGAGCGGAGAGATGCGCCCTTAC 1012  
DB 245 TGGAAATGATGAAGGCGAGACCGTCAAGAGCGGAGAGATGCGCCCTTAC 194

RESULT 3  
ADQ86032  
ID ADQ86032 standard; cDNA; 2372 BP.  
XX  
AC ADQ86032:  
XX  
DT 07-OCT-2004 (first entry)  
XX  
DE Human tumour-associated antigenic target (TAT) cDNA sequence #2904.  
XX  
KW human; tumour-associated antigenic target; TAT; cytostatic; gene therapy;

KW cancer; cell proliferative disorder; gene; ss.  
XX Homo sapiens.  
OS  
PN MO2004060270-A2.  
XX  
XX 22-JUL-2004.  
PD  
XX  
PF 15-OCT-2003; 2003MO-US029126.  
XX  
PR 18-OCT-2002; 2002US-0418988P.  
XX  
XX (GENT) GENENTECH INC.  
PA (WU/D) WU T D.  
PA (ZHOU) ZHOU Y.  
XX  
PI Wu TD, Zhou Y;  
XX  
DR WPI; 2004-534300/51.  
XX  
PT New nucleic acid molecule and encoded polypeptide, for diagnosing,  
PT preventing or treating cell proliferative disorders such as cancer.  
XX  
PS Claim 1; SEQ ID NO 2904; 5504bp; English.  
XX  
CC The present invention describes an isolated tumour-associated antigenic  
CC target (TAT) nucleic acid comprising: (a) any of 4622 nucleotide  
CC sequences (see SEQ ID NO:1 to 4622); (b) the full-length coding region of  
CC (a); (c) the complement of (a) or (b); (d) a sequence that has 80%  
CC sequence identity to (a)-(c); or (e) a sequence that hybridises to (a)-  
CC (c). Also described: (1) an expression vector comprising the above  
CC nucleic acid; (2) a host cell comprising the above expression vector; (3)  
CC a process for producing a polypeptide; (4) an isolated polypeptide  
CC comprising: (a) an amino acid sequence encoded by any of the above  
CC nucleotide sequences; (b) an amino acid sequence encoded by the full-  
CC length coding region of the above nucleotide sequences; or (c) a sequence  
CC having at least 80% identical to (a) or (b); (5) a chimeric polypeptide  
CC comprising the above polypeptide fused to a heterologous polypeptide; (6)  
CC an isolated antibody that binds to the above polypeptide; (7) a process  
CC for producing the antibody; (8) an isolated oligopeptide that binds to  
CC the above polypeptide; (9) a tumour-associated antigenic target (TAT)  
CC binding organic molecule that binds to the above polypeptide; (10) a  
CC composition of matter comprising the above (chimeric) polypeptide,  
CC antibody, oligopeptide or TAT binding organic molecule, in combination  
CC with a carrier; (11) an article of manufacture comprising a container and  
CC the composition of matter contained within the container; (12) methods of  
CC inhibiting the growth of a cell that expresses the above protein, where  
CC the growth of the cell is at least in part dependent upon a growth  
CC potentiating effect of the above protein; (13) a method of  
CC therapeutically treating a mammal having a cancerous tumour comprising  
CC cells that express the above protein; (14) a method of determining the  
CC presence of a protein in a sample suspected of containing the protein  
CC described above; (15) methods of diagnosing the presence of a tumour in a  
CC mammal; (16) a method for treating or preventing a cell proliferative  
CC disorder associated with increased expression or activity of the above  
CC protein; and (17) a method of binding an antibody, oligopeptide or  
CC organic molecule to a cell that expresses the protein described above.  
CC The TAT sequences have cytostatic activities, and can be used in gene  
CC therapy. The composition and methods are useful for diagnosing,  
CC preventing or treating cancer. The composition is also used for preparing  
CC a medicament for the therapeutic treatment or diagnostic detection of a  
CC cell proliferative disorder or cancer. The present sequence represents a  
CC human TAT cDNA sequence from the present invention.  
XX  
SQ Sequence 2372 BP; 628 A; 589 C; 668 G; 487 T; 0 U; 0 Other;

Query Match 100.0%; Score 1012; DB 13; Length 2372;  
Best Local Similarity 100.0%; Pred. No. 3.1e-233;  
Matches 1012; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGGACCGGTGATGAGGAGCGCTGAGAGCGAGGCTGCTCAATAGCGCTGGAGAC 60  
DB 1000 AGGACCGGTGATGAGGAGCGCTGAGAGCGAGGCTGCTCAATAGCGCTGGAGAC 1059

QY 61 CATGATGAAGGACGAGATTTGATCTAGGCCCCAGCTGTAGATCCAGATGAAGACTCTGA 120  
 DB 1060 CATGATGAAGGACGAGATTTGATCTAGGCCCCAGCTGTAGATCCAGATGAAGACTCTGA 1119  
 QY 121 CAACAGTGAATTTATGTAACAAGATTAATGAAGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 180  
 DB 1120 CAACAGTGAATTTATGTAACAAGATTAATGAAGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 1179  
 QY 181 CTTCTTTAAGCAGTGTGGGTTGTTAAGATGAACAAGAGAACTGGGCAACCTATGATCCA 240  
 DB 1180 CTTCTTTAAGCAGTGTGGGTTGTTAAGATGAACAAGAGAACTGGGCAACCTATGATCCA 1239  
 QY 241 CATCTACCTGGAGCAAGGAAACAGGAAAGCCCAAGGAGTCCAGACAGTGTCTATGAAGA 300  
 DB 1240 CATCTACCTGGAGCAAGGAAACAGGAAAGCCCAAGGAGTCCAGACAGTGTCTATGAAGA 1299  
 QY 301 CCCACCACTGCAAGGCTGCTGGAATGTTGATGAGAAAGATTTCAAGGAGCA 360  
 DB 1300 CCCACCACTGCAAGGCTGCTGGAATGTTGATGAGAAAGATTTCAAGGAGCA 1359  
 QY 361 ACTTAAAGTCTCTGCTGGAAGAAAGCTTCAATGAAGATGCGGGTGTCTGCC 420  
 DB 1360 ACTTAAAGTCTCTGCTGGAAGAAAGCTTCAATGAAGATGCGGGTGTCTGCC 1419  
 QY 421 ACCCGGTGAGGAGGAGGAGTGCACACCACTCCGTGAGGTCCAGAGGCCAGAGG 480  
 DB 1420 ACCCGGTGAGGAGGAGGAGTGCACACCACTCCGTGAGGTCCAGAGGCCAGAGG 1479  
 QY 481 TCCGTGGGAGCCCATGAGGTGCAATGAGAGGCGGTGAGAGATGAGAGGCTTCCCTCC 540  
 DB 1480 TCCGTGGGAGCCCATGAGGTGCAATGAGAGGCGGTGAGAGATGAGAGGCTTCCCTCC 1539  
 QY 541 AAGAGAACCCCGGGGTTCCCGAGGAAACCCCTCTGAGAGAGAAAGTCCAGACCGAGC 600  
 DB 1540 AAGAGAACCCCGGGGTTCCCGAGGAAACCCCTCTGAGAGAGAAAGTCCAGACCGAGC 1599  
 QY 601 TGGAGACTGCAAGTCCCAATCCGGGTGTGAGAAACCAACACTTCCCTGAGAAACAGA 660  
 DB 1600 TGGAGACTGCAAGTCCCAATCCGGGTGTGAGAAACCAACACTTCCCTGAGAAACAGA 1659  
 QY 661 GTGCAACAGTGTAAAGCCCAAAAGCTGAAAGCTTCCCGCCACCTTTCGCGCCCC 720  
 DB 1660 GTGCAACAGTGTAAAGCCCAAAAGCTGAAAGCTTCCCGCCACCTTTCGCGCCCC 1719  
 QY 721 GGGTGTGATCTGTGAGAGAGTGGCCCTGTGTCATGCGGGAGAGAAAGAGTGGCTCAT 780  
 DB 1720 GGGTGTGATCTGTGAGAGAGTGGCCCTGTGTCATGCGGGAGAGAAAGAGTGGCTCAT 1779  
 QY 781 GGAATCGTGTGTCCTCGGTGGAATGTTCAAGAGTGGCCGTGTGAGAGACAGAGTGGCTT 840  
 DB 1780 GGAATCGTGTGTCCTCGGTGGAATGTTCAAGAGTGGCCGTGTGAGAGACAGAGTGGCTT 1839  
 QY 841 CCGTGTGAGCCGGGGCAATGAGCCGAGGTGCTTGTGTGAGAGAAAGAGTGGCCCTGAG 900  
 DB 1840 CCGTGTGAGCCGGGGCAATGAGCCGAGGTGCTTGTGTGAGAGAAAGAGTGGCCCTGAG 1899  
 QY 901 GGGGCCCCCTTGTGATGGAACAGATGGAGAGAAAGAGAGAGAGAGAGAGAGAGAGAGAG 960  
 DB 1900 GGGGCCCCCTTGTGATGGAACAGATGGAGAGAAAGAGAGAGAGAGAGAGAGAGAGAGAG 1959  
 QY 961 TGGAAAAATGATTAAG 1012  
 DB 1960 TGGAAAAATGATTAAG 2011

## RESULT 4

ID AD516277 Standard; cDNA; 2390 BP.  
 AC AD516277;  
 XX  
 DT 18-NOV-2004 (first entry)

XX DE Human cDNA encoding the Ewing sarcoma protein SegID 1.  
 XX KW human; ss; gene; Ewing sarcoma; EWS; prostatic cancer; alopecia; acne;  
 XX KW hypogonadism; androgen-resistance syndrome; testicular feminisation.  
 XX OS Homo sapiens.  
 XX PH Key Location/Qualifiers  
 XX FT CDS 44..2014  
 XX FT /tag= a  
 XX FT /product= "Ewing sarcoma protein"  
 XX PN EPI45190-A1.  
 XX PD 08-SEP-2004.  
 XX PF 16-FEB-2004; 2004EP-00003422.  
 XX PR 04-MAR-2003; 2003DE-01009280.  
 XX PR 25-APR-2003; 2003US-0465692P.  
 XX PA (SCHD) SCHERING AG.  
 XX PI Obendorf M, Wolf S;  
 XX PI WPI, 2004-627861/61.  
 XX DR P-PSDB; AD516278.  
 XX PT Determining the hormonal effects of substances, used to identify  
 PT pharmaceutical, e.g. for treatment of androgen receptor dysfunction,  
 PT from modulating interaction between nuclear receptors and Ewing sarcoma  
 PT protein.  
 XX PS Claim 7, SEQ ID NO 1, 30pp; German.  
 XX CC This invention relates to a novel modulators that alter the interaction  
 CC between the Ewing sarcoma protein (EWS) and its nuclear receptor, as well  
 CC as the screening method thereof. Specifically, it refers to determining  
 CC and identifying a hormonal effect brought about by test compounds that  
 CC modulate either the binding of EWS to the nuclear receptor or the ligand-  
 CC induced activity of this receptor. The present invention describes the  
 CC nuclear receptors as including oestrogen, progesterone, thyroid hormone,  
 CC vitamin D, and retinoic acid receptors, most preferably they are androgen  
 CC receptors. Accordingly, these modulators may be used in the development  
 CC of pharmaceutical compositions that can diagnose and be used to treat  
 CC diseases associated with receptor dysfunction such as prostatic cancer,  
 CC alopecia, acne, hypogonadism and androgen-resistance syndrome e.g.  
 CC testicular feminisation. This method provides reliable, sensitive,  
 CC simple, inexpensive and rapid assessment of the hormonal effects of these  
 CC test compounds. This polynucleotide sequence is the cDNA encoding the  
 CC human Ewing sarcoma protein of the invention.  
 XX SQ Sequence 2390 BP; 646 A; 589 C; 668 G; 487 T; 0 U; 0 Other;  
 XX  
 Query Match 100.0%; Score 1012; DB 13; Length 2390;  
 Best Local Similarity 100.0%; Pred. No. 3, 1e-233;  
 Matches 1012; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 AGGACCCGGTGAATGAGAGAGCGCTGAGAGCGAGGTGCTTCAATGAAGCTGTGAGACC 60  
 DB 1000 AGGACCCGGTGAATGAGAGAGCGCTGAGAGCGAGGTGCTTCAATGAAGCTGTGAGACC 1059  
 QY 61 CATGATGAAGGACGAGATTTGATCTAGGCCCCAGCTGTAGATCCAGATGAAGACTCTGA 120  
 DB 1060 CATGATGAAGGACGAGATTTGATCTAGGCCCCAGCTGTAGATCCAGATGAAGACTCTGA 1119  
 QY 121 CAACAGTGAATTTATGTAACAAGATTAATGAAGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 180  
 DB 1120 CAACAGTGAATTTATGTAACAAGATTAATGAAGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 1179  
 QY CTTCTTTAAGCAGTGTGGGTTGTTAAGATGAACAAGAGAACTGGGCAACCTATGATCCA 240



Db 1180 CTTCTTAAGCAGTGTGGGTTGTTAAGATGAACAAGAACTGGCAACCATGATCCA 1239  
 QY 241 CATCTACCTGGAACAAGAAACAGAAAGCCCAAGGAGATGCCAGATGTCCTATGAAGA 300  
 Db 1240 CATCTACCTGGAACAAGAAACAGAAAGCCCAAGGAGATGCCAGATGTCCTATGAAGA 1299  
 QY 301 CCCAACCACTGCGCAAGGCTCGGTGAATGTTGATGGGAAAGATTTTCAAGGAGACAA 360  
 Db 1300 CCCAACCACTGCGCAAGGCTCGGTGAATGTTGATGGGAAAGATTTTCAAGGAGACAA 1359  
 QY 361 ACTTAAAGTCTCCCTTCTCGAAGAACCTTCAATGAACATATGCGGGGTGGTCTGCC 420  
 Db 1360 ACTTAAAGTCTCCCTTCTCGAAGAACCTTCAATGAACATATGCGGGGTGGTCTGCC 1419  
 QY 421 ACCCCGTGAGGGGCAAGAGGATGCCACCACTCCGTGAGAGTCCAGAGAGCCCAAGAGG 480  
 Db 1420 ACCCCGTGAGGGGCAAGAGGATGCCACCACTCCGTGAGAGTCCAGAGAGCCCAAGAGG 1479  
 QY 481 TCCTGGGGGACCAATGGGTGGCATGGAGGCGGTGAGAGAGATGAGAGAGGCTTCCCTCC 540  
 Db 1480 TCCTGGGGGACCAATGGGTGGCATGGAGGCGGTGAGAGAGATGAGAGAGGCTTCCCTCC 1539  
 QY 541 AAGAGACCCCGGGGTTCCCGAGGAAACCCCTCTGAGAGAGAAACGTCCAGCAGCAGC 600  
 Db 1540 AAGAGACCCCGGGGTTCCCGAGGAAACCCCTCTGAGAGAGAAACGTCCAGCAGCAGC 1599  
 QY 601 TGGAGACTGGCAGTGTCCCAATCCGGGTGTGAAACAGAACTTCCCTGAGAGACAGA 660  
 Db 1600 TGGAGACTGGCAGTGTCCCAATCCGGGTGTGAAACAGAACTTCCCTGAGAGACAGA 1659  
 QY 661 GTGCAACCACTGTAAAGGCCCCCAAAAGCCTGAAGGCTTCCCGCCACCTTCCGCCCC 720  
 Db 1660 GTGCAACCACTGTAAAGGCCCCCAAAAGCCTGAAGGCTTCCCGCCACCTTCCGCCCC 1719  
 QY 721 GGGTGTGATCTGTGACAGAGTGGCCTGTGTGCATCCGGGAGAGAAAGTGGCTCAT 780  
 Db 1720 GGGTGTGATCTGTGACAGAGTGGCCTGTGTGCATCCGGGAGAGAAAGTGGCTCAT 1779  
 QY 781 GGATCTGTGTGTCCCGGTGAATGTTCAAGAGTGGCCGTGTGAGACAGAGGTGGCTT 840  
 Db 1780 GGATCTGTGTGTCCCGGTGAATGTTCAAGAGTGGCCGTGTGAGACAGAGGTGGCTT 1839  
 QY 841 CCGTGTGCGCGGGGCAATGACCGAGGTGGCTTGTGTGAGAGAAAGAGTGGCTTGG 900  
 Db 1840 CCGTGTGCGCGGGGCAATGACCGAGGTGGCTTGTGTGAGAGAAAGAGTGGCTTGG 1899  
 QY 901 GGGGCCCCCTGGAACCTTTGATGGAACAGATGGGAGAAAGAGAGAGTGGAGGACC 960  
 Db 1900 GGGGCCCCCTGGAACCTTTGATGGAACAGATGGGAGAAAGAGAGAGTGGAGGACC 1959  
 QY 961 TGGAAAAATGGAATAAAGCGAGACCGTCAAGAGCGCGAGAGATCGGCTTAC 1012  
 Db 1960 TGGAAAAATGGAATAAAGCGAGACCGTCAAGAGCGCGAGAGATCGGCTTAC 2011  
 RESULT 5  
 AAQ50643  
 ID AAQ50643 standard; cDNA; 2371 BP.  
 XX  
 AC AAQ50643;  
 XX  
 DT 25-MAR-2003 (revised)  
 DT 26-MAY-1994 (first entry)  
 XX  
 DE Human Ews gene clone Bf1AC5 from foetal brain cDNA library.  
 XX  
 KW chromosomal translocation; chimeric; chimeric; Ewing sarcoma; Ews gene;  
 KW malignant melanoma; hum-fli-1;  
 KW primitive peripheral neuroectodermal tumour; human chromosome 11;  
 KW human chromosome 22; ss.  
 XX  
 OS Homo sapiens.  
 XX

FH Key Location/Qualifiers  
 FT CDS 25..1995  
 FT /tag= a  
 FT /product= "EWS protein"  
 FT /transl\_except= pos:1729..1731; aa:Val  
 FT polyA\_signal 2143..2148  
 FT /tag= b  
 FT polyA\_signal 2311..2336  
 FT /tag= c  
 XX  
 XX MO9323549-A2.  
 XX  
 XX 25-NOV-1993.  
 XX  
 XX 19-MAY-1993; 93MO-FR000494.  
 XX  
 XX 20-MAY-1992; 92FR-00006123.  
 XX  
 XX (CNRS ) CNRS CENT NAT RECH SCI.  
 XX  
 XX Aurias A, Delattre O, Desmaza C, Meiot T, Peter M, Plougastel B,  
 PI Thomas G, Zucman J;  
 XX  
 DR WPI: 1993-386580/48.  
 DR P-PSDB; AAR44555.  
 XX  
 XX New nucleic acid of EWS gene and its hybrid(s) - contg. gene sequence  
 PT involved in chromosomal translocation, also derived mRNA, probes, fusion  
 PT proteins etc., for diagnosis and treatment of Ewing sarcoma and melanoma.  
 XX  
 PS Disclosure; Fig 6; 123p; French.  
 CC  
 CC The probes 22R3 and 22R12 were used to screen a human foetal brain cDNA  
 CC library (Stratagene cat. # 936206). The clone Bf1AC5 was identified and  
 CC sequenced. It represents the entire coding region and 3'-UTR of the Ews  
 CC gene. (Updated on 25-MAR-2003 to correct PN field.)  
 XX  
 SQ Sequence 2371 BP; 639 A; 587 C; 659 G; 486 T; 0 U; 0 Other;  
 Query Match 99.8%; Score 1010.4; DB 2; Length 2371;  
 Best Local Similarity 99.9%; Pred. No. 7.5e-233;  
 Matches 1011; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 AGGAGCGGTGAGTAAAGGCGAGCGCTGAGAGCGAGGTGGCTTCAATAGCCTGTGACC 60  
 Db 981 AGGAGCGGTGAGTAAAGGCGAGCGCTGAGAGCGAGGTGGCTTCAATAGCCTGTGACC 1040  
 QY 61 CATGATGAAGACCAAGATCTTATCTAGAGCCCACTGTATGATCCAGATGAACATCTGA 120  
 Db 1041 CATGATGAAGACCAAGATCTTATCTAGAGCCCACTGTATGATCCAGATGAACATCTGA 1100  
 QY 121 CAACAGTCAATTATGTAACAAGATTAATGAAGTGTGAAGTCTGATGATGTGGCAGA 180  
 Db 1101 CAACAGTCAATTATGTAACAAGATTAATGAAGTGTGAAGTCTGATGATGTGGCAGA 1160  
 QY 181 CTTCTTAAAGCAGTGTGGGTTGTTAAGATGAACAAGAACTGGCAACCCATGATCCA 240  
 Db 1161 CTTCTTAAAGCAGTGTGGGTTGTTAAGATGAACAAGAACTGGCAACCCATGATCCA 1220  
 QY 241 CATCTACCTGGAACAAGAAACAGAAAGCCCAAGGAGATGCCAGATGTCCTATGAAGA 300  
 Db 1221 CATCTACCTGGAACAAGAAACAGAAAGCCCAAGGAGATGCCAGATGTCCTATGAAGA 1280  
 QY 301 CCCAACCACTGCGCAAGGCTCGGTGAATGTTGATGGGAAAGATTTTCAAGGAGACAA 360  
 Db 1281 CCCAACCACTGCGCAAGGCTCGGTGAATGTTGATGGGAAAGATTTTCAAGGAGACAA 1340  
 QY 361 ACTTAAAGTCTCCCTTCTCGAAGAACCTTCAATGAACATATGCGGGGTGGTCTGCC 420  
 Db 1341 ACTTAAAGTCTCCCTTCTCGAAGAACCTTCAATGAACATATGCGGGGTGGTCTGCC 1400  
 QY 421 ACCCCGTGAGGGGCAAGAGGATGCCACCACTCCGTGAGAGTCCAGAGAGCCCAAGAGG 480

D	b		1401	ACCCCGTAGGGGCAAGAGCATGCACCAACCACCTCCGTGAGAGTCCAGAAGGCCAAGAGG	1460
O	y		481	TCCCTGGGGGACCCATGGGTCGCATGGAGGCGCGTGAGAGAGATAGAGAGGCTTCCCCTCC	540
D	b		1461	TCCTGGGGGGAACCATGGGTTGGCATTGGGAGGCGCGTGAGAGAGATAGAGAGGCTTCCCCTCC	1520
O	y		541	AAGAGAACCCCCGGGGTTTCCCGAGGAAACCCCTCTGAGAGAGAGAAAGTCCAGCACCGAAGC	600
D	b		1521	AAGAGAACCCCCGGGGTTTCCCGAGGAAACCCCTCTGAGAGAGAGAAAGTCCAGCACCGAAGC	1580
O	y		601	TGGAAGACTGGAGAGTGTGCCCAATCCGGGTTGTGAAAACAAGAACTTGGCTTGGAGAACAGA	660
D	b		1581	TGGAAGACTGGAGAGTGTGCCCAATCCGGGTTGTGAAAACAAGAACTTGGCTTGGAGAACAGA	1640
O	y		661	GTCGACCCAGTGTAAAGGCCCCCAAAGCCCTGAAGGCTTCTCCCGCACCCCTTTCGCCCCC	720
D	b		1641	GTCGACCCAGTGTAAAGGCCCCCAAAGCCCTGAAGGCTTCTCCCGCACCCCTTTCGCCCCC	1700
O	y		721	GGGTGTGATCGTGGCAGAGGTGGCCCTGTGTGGCAATGCGGGGAGAGAGAGTGGCTTCAT	780
D	b		1701	GGGTGTGATCGTGGCAGAGGTGGCCCTGTGTGGCAATGCGGGGAGAGAGAGTGGCTTCAT	1760
O	y		781	GGATCGTGTGTGTCCCGGTGGAATGTTGAGAGGTGGCCGCTGTGGAGACAGAGGTGGCTT	840
D	b		1761	GGATCGTGTGTGTCCCGGTGGAATGTTGAGAGGTGGCCGCTGTGGAGACAGAGGTGGCTT	1820
O	y		841	CCGTGTGTGGCCGGGCGCATGGAACCGAGGTGGCTTTGTGTGAGAGAAACGAGGTGGCCCTGG	900
D	b		1821	CCGTGTGTGGCCGGGCGCATGGAACCGAGGTGGCTTTGTGTGAGAGAAACGAGGTGGCCCTGG	1880
O	y		901	GGGGCCCCCTTGGA CTTTGTATGAAACAGATGGAGAGAAAGAGAGAGACGTGAGAGACC	960
D	b		1881	GGGGCCCCCTTGGA CTTTGTATGAAACAGATGGAGAGAAAGAGAGAGACGTGAGAGACC	1940
O	y		961	TGGAAAAATGGATTAAAGCGAGCAACCTCAAGAGGCGAGAGATGGGCCCTTAC	1012
D	b		1941	TGGAAAAATGGATTAAAGCGAGCAACCTCAAGAGGCGAGAGATGGGCCCTTAC	1992
<hr/>					
RESULT 6					
AAS70647					
ID	AAS70647 standard; cDNA, 2388 BP.				
XX					
AC	AAS70647;				
XX					
DT	13-FEB-2002 (first entry)				
XX					
DE	DNA encoding novel human diagnostic protein #6451.				
XX					
KW	Human; chromosome mapping; gene mapping; gene therapy; forensic;				
XX	Food supplement; medical imaging; diagnostic; disorder; ss.				
OS	Homo sapiens.				
XX					
PN	WO200175067-A2.				
XX					
PD	11-OCT-2001.				
XX					
PF	30-MAR-2001; 2001MO-US008631.				
XX					
PR	31-MAR-2000; 2000US-00540217.				
XX					
PR	23-AUG-2000; 2000US-00649167.				
PA	(HYSE-) HYSEQ INC.				
XX					
PI	Dymanac RT, Liu C, Tang YT;				
XX					
DR	WPI; 2001-639362/73.				
XX					
DR	P-PSDB; ABG06460.				
PT	New isolated polynucleotide and encoded polypeptides, useful in				
PT	diagnostics, forensics, gene mapping, identification of mutations				
PT	responsible for genetic disorders or other traits and to assess				

Query Match	Best Local Similarity	99.8%;	Score 1010.4;	DB 5;	Length 2388;	
Matches 1011;	Conservative	0;	Mismatches 1;	Indels	0;	Gaps 0;
QY	1	AGGACGGCGGTGAAATGAGGACGCGTGGAGAGCGAGGTGGCTTCATTAAGCTGTGTGAC	60			
DB	1017	AGGACGGCGGTGAAATGAGGACGCGTGGAGAGCGAGGTGGCTTCATTAAGCTGTGTGAC	1076			
QY	61	CATGATGATGAGACCAACATCTTATCTTAGGCCCACTGTGATTCACATGAAAGCTCGA	120			
DB	1077	CATGATGATGAGACCAACATCTTATCTTAGGCCCTCTGTGATTCACATGAAAGCTCGA	1136			
QY	121	CACAGTGCATATTATGTACAAAGATTAAATGACAGTGTGACTTAATGATCTGTGGAGA	180			
DB	1137	CACAGTGCATATTATGTACAAAGATTAAATGACAGTGTGACTTAATGATCTGTGGAGA	1196			
QY	181	CTTCTTTAAGCAGTGTGGGCTTTGTAAGATGAACAAGAACTGGGCAACCCATGATCCA	240			
DB	1197	CTTCTTTAAGCAGTGTGGGCTTTGTAAGATGAACAAGAACTGGGCAACCCATGATCCA	1256			
QY	241	CATCTACTGGACAAGAAACAGAAAGCCAAAGGGGATGCCACATGTCTTAATGAA	300			
DB	1257	CATCTACTGGACAAGAAACAGAAAGCCAAAGGGGATGCCACATGTCTTAATGAA	1316			
QY	301	CCGACCCACATGACCAAGGCTGCCGCGGAAATGATTGATGGGAAATTTTCAAGGGAGCA	360			
DB	1317	CCGACCCACATGACCAAGGCTGCCGCGGAAATGATTGATGGGAAATTTTCAAGGGAGCA	1376			
QY	361	ACTTAAATCTCCCTTGTCCGAGAAAGCTTCCATGAACAGTATGCGGGGTGTCTGCC	420			
DB	1377	ACTTAAATCTCCCTTGTCCGAGAAAGCTTCCATGAACAGTATGCGGGGTGTCTGCC	1436			
QY	421	ACCCGATGAGGACAGAGGATGCCACCACTCCGTGAGAGTCCAGAGGCCCGAGAGG	480			
DB	1437	ACCCGATGAGGACAGAGGATGCCACCACTCCGTGAGAGTCCAGAGGCCCGAGAGG	1496			
QY	481	TTCCTGGGGGACCAATGGGTGCGCATGGGAGGCGGTGAGAGATGAGGAGGCTTCCCTCC	540			
DB	1497	TTCCTGGGGGACCAATGGGTGCGCATGGGAGGCGGTGAGAGATGAGGAGGCTTCCCTCC	1556			
QY	541	AAGAGGACCCCGGGGTTCCTCGAGGGAACCTCTGTGAGAGGAAACGTCCAGACCGAGC	600			
DB	1557	AAGAGGACCCCGGGGTTCCTCGAGGGAACCTCTGTGAGAGGAAACGTCCAGACCGAGC	1616			
QY	601	TGGAACTGCGAGTGTCCCAATCCGGGTGTGAGAAACGAACTTCGCTGAGAACAGA	660			

Db 1617 TGAAGACTGCGAGTGTCCCAATCCGGTGTGTGAAAACAACCTTCCCTGGAGAACAGA 1676  
 Qy 661 GTGCAACCACTGTAAAGAGCCCAAAAGCCTGAAGCTTCTCCCGCACCTTTCCGCCCC 720  
 Db 1677 GTCAACCACTGTAAAGAGCCCAAAAGCCTGAAGCTTCTCCCGCACCTTTCCGCCCC 1736  
 Qy 721 GGGTGTGATCTGTGCGAGAGTGGCCCTGTGTGCAATCCGGGAGGAAGAGTGGCTCAT 780  
 Db 1737 GGGTGTGATCTGTGCGAGAGTGGCCCTGTGTGCAATCCGGGAGGAAGAGTGGCTCAT 1796  
 Qy 781 GATTCGTGTGTCTCCCGGTGAATGTTCAGAGTGGCCGTGTGTGAGACAGAGTGGCTT 840  
 Db 1797 GATTCGTGTGTCTCCCGGTGAATGTTCAGAGTGGCCGTGTGTGAGACAGAGTGGCTT 1856  
 Qy 841 CCGTGTGTGCGCGGAGCAGTGAACCGAGGTGGCTTGTGTGAGAGAACAGAGTGGCCCTG 900  
 Db 1857 CCGTGTGTGCGCGGAGCAGTGAACCGAGGTGGCTTGTGTGAGAGAACAGAGTGGCCCTG 1916  
 Qy 901 GGGGCCCCCTTGACCTTTGATGAAACAGATGGAGAGGAAGAGAGACGTGGAGACC 960  
 Db 1917 GGGGCCCCCTTGACCTTTGATGAAACAGATGGAGAGGAAGAGAGACGTGGAGACC 1976  
 Qy 961 TGGAAAAATGATTAAGAGCAGACCGTCAAGAGCGCAGAGATCGGCTTAC 1012  
 Db 1977 TGGAAAAATGATTAAGAGCAGACCGTCAAGAGCGCAGAGATCGGCTTAC 2028  
 RESULT 7  
 ABK84628  
 ID ABK84628 standard; cDNA, 2390 BP.  
 XX ABK84628;  
 DT 14-AUG-2002 (first entry)  
 DE Human cDNA differentially expressed in granulocytic cells #1199.  
 XX  
 KW Human; ss; granulocytic cell; DNA chip; bacterial infection;  
 KW viral infection; parasitic infection; protozoal infection;  
 KW fungal infection; sterile inflammatory disease; psoriasis;  
 KW rheumatoid arthritis; glomerulonephritis; asthma; thrombosis;  
 KW cardiac reperfusion injury; renal reperfusion injury; ARDS;  
 KW adult respiratory distress syndrome; inflammatory bowel disease;  
 KW Crohn's disease; ulcerative colitis; peridontal disease;  
 KW granulocyte activation; chronic inflammation; allergy.  
 XX  
 OS Homo sapiens.  
 XX  
 PN MO200228999-A2.  
 XX  
 PD 11-APR-2002.  
 XX  
 PF 03-OCT-2001; 2001MO-US030821.  
 XX  
 PR 03-OCT-2000; 2000US-0237189P.  
 XX  
 PA (GENE-) GENE LOGIC INC.  
 XX  
 PI Beazer-Barclay Y, Weissman SM, Yamaga S, Vockley J;  
 XX WPI, 2002-435328/46.  
 DR  
 XX  
 PT Detecting granulocyte activation by detecting differential expression of  
 PT genes associated with granulocyte activation, which serves as diagnostic  
 PT markers that is useful for monitoring disease states and drug toxicity.  
 XX  
 PS Claim 1; SEQ ID NO 1199; 114bp; English.  
 XX  
 CC The invention relates to detecting (M1) granulocyte (GC) activation  
 CC (GCA), by detecting the level of expression of gene(s) (Gs) identified by  
 CC DNA chip analysis as given in the specification, and comparing the  
 CC expression level to an expression level in an unactivated GC, where

CC differential expression of Gs is indicative of GCA. Also included are  
 CC modulating (M2) Gs by contacting GC with an agent that alters the  
 CC expression of at least one gene in Gs; (2) screening (M3) for an agent  
 CC capable of modulating GCA or an inflammation (especially chronic) in a  
 CC tissue, an allergic response in a subject, exposure of a subject to a  
 CC pathogen or sterile inflammatory disease using the gene expression  
 CC profile; (3) detecting (M4) an inflammation (especially chronic) in a  
 CC tissue, an allergic response in a subject, exposure of a subject to a  
 CC pathogen or sterile inflammatory disease, by detecting the level of  
 CC expression in a sample of the tissue of gene(s) from Gs, where the level  
 CC of expression of the gene is indicative of inflammation; (4) treating  
 CC (M5) an inflammation (especially chronic) or in a tissue, an allergic  
 CC response in a subject, exposure of a subject to a pathogen or sterile  
 CC inflammatory disease, by contacting a tissue having inflammation with an  
 CC agent that modulates the expression of gene(s) from Gs in the tissue. M1  
 CC is useful for detecting GCA; M2 is useful for modulating Gs; M3 is useful  
 CC for screening an agent capable of modulating GCA preferably in an  
 CC inflammation in a tissue; M4 is useful for detecting an inflammation  
 CC (especially chronic) in a tissue, an allergic response in a subject,  
 CC exposure of a subject to a pathogen or sterile inflammatory disease (e.g.  
 CC psoriasis, rheumatoid arthritis, glomerulonephritis, asthma, thrombosis,  
 CC cardiac reperfusion injury, renal reperfusion injury, ARDS, adult  
 CC respiratory distress syndrome, inflammatory bowel disease, Crohn's  
 CC disease, ulcerative colitis, peridontal disease; also bacterial  
 CC infection, viral infection, parasitic infection, protozoal infection,  
 CC fungal infection and M5 is useful for treating one of the above  
 CC conditions. The present sequence represents a gene differentially  
 CC expressed in granulocytes. Note: The sequence data for this patent did  
 CC not form part of the printed specification, but was obtained in  
 CC electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences  
 XX  
 SQ Sequence 2390 BP; 645 A; 589 C; 668 G; 488 T; 0 U; 0 Other;  
 Query Match 99.8%; Score 1010.4; DB 6; Length 2390;  
 Best Local Similarity 99.9%; Pred. No. 7.5e-233;  
 Matches 1011; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 Qy 1 AGGAGCGGTGGTGAATGGGAGCGCTGAGAGCGAGGTGGCTTCAATAGCCTGTGGACC 60  
 Db 1000 AGGAGCGGTGGTGAATGGGAGCGCTGAGAGCGAGGTGGCTTCAATAGCCTGTGGACC 1059  
 Qy 61 CATGATGAAGAAGCAAGATCTTATTAAGGCCCACTGTATGATCCAGATGAACCTTGA 120  
 Db 1060 CATGATGAAGAAGCAAGATCTTATTAAGGCCCTCTGTATGATCCAGATGAACCTTGA 1119  
 Qy 121 CAACAGTGAATTATGTAACAAGATTAAATGACAGTGTACTCTGATGATCTGGCAGA 180  
 Db 1120 CAACAGTGAATTATGTAACAAGATTAAATGACAGTGTACTCTGATGATCTGGCAGA 1179  
 Qy 181 CTTCTTTAAGCAGTGTGGGTTGTAAATGAACAAGAACTGGCAACCCATGATCCA 240  
 Db 1180 CTTCTTTAAGCAGTGTGGGTTGTAAATGAACAAGAACTGGCAACCCATGATCCA 1239  
 Qy 241 CATCTACCTGAGACAGAGAAACAGAAAGCCCAAGGCGATGCCAGTGTCTTATGAAGA 300  
 Db 1240 CATCTACCTGAGACAGAGAAACAGAAAGCCCAAGGCGATGCCAGTGTCTTATGAAGA 1299  
 Qy 301 CCCACCCATGCGCAAGGCTGCGGTGAATGTGTGATGGAGAAAGATTTTCAAGGAGCAA 360  
 Db 1300 CCCACCCATGCGCAAGGCTGCGGTGAATGTGTGATGGAGAAAGATTTTCAAGGAGCAA 1359  
 Qy 361 ACTTAAAGTCTCCCTTGTCTGGAAGAAGCTCCCAATGAACATGATCGGGGTGTGCC 420  
 Db 1360 ACTTAAAGTCTCCCTTGTCTGGAAGAAGCTCCCAATGAACATGATCGGGGTGTGCC 1419  
 Qy 421 ACCCGCTGAGGAGAGAGCAATGCCACCACTCCGTGTGAGGTCCAGAGGCCCAAGAG 480  
 Db 1420 ACCCGCTGAGGAGAGAGCAATGCCACCACTCCGTGTGAGGTCCAGAGGCCCAAGAG 1479  
 Qy 481 TCTGTGGGAGACCATATGATGTGCAATGAGGAGCGGTGAGAGAGATGAGAGGCTTCCCTCC 540  
 Db 1480 TCTGTGGGAGACCATATGATGTGCAATGAGGAGCGGTGAGAGAGATGAGAGGCTTCCCTCC 1539

OY	541	AAAGAGACCCCGGGGTTCCCGAGGAAACCCCTCTGAGAGAGAAACGTCACGACCCAGC	600
Db	1540	AAAGAGACCCCGGGGTTCCCGAGGAAACCCCTCTGAGAGAGAAACGTCACGACCCAGC	1539
OY	601	TGAGACTGGCAGTGTCCCAATCCGGGTTGTGAAACAGAACTTCGCTTGAGAAACAG	660
Db	1600	TGAGACTGGCAGTGTCCCAATCCGGGTTGTGAAACAGAACTTCGCTTGAGAAACAG	1659
OY	661	GTGACACCAAGTGTAAAGGCCCAAAAGCCGAAAGGCTTCCTCCCGCACCCCTTCCGCCCC	720
Db	1660	GTGACACCAAGTGTAAAGGCCCAAAAGCCGAAAGGCTTCCTCCCGCACCCCTTCCGCCCC	1719
OY	721	GCGTGTGATCGTGGCAGAGGTGGCCCTGTGTGCATTCGAGGAGAGAAAGGTGGCTCAT	780
Db	1720	GCGTGTGATCGTGGCAGAGGTGGCCCTGTGTGCATTCGAGGAGAGAAAGGTGGCTCAT	1779
OY	781	GGAATGTGTGTGGTCCCCGGTGGAAATTTCAAGAGTGGCCGTGTGTGAACAAGTGTGTT	840
Db	1780	GGAATGTGTGTGGTCCCCGGTGGAAATTTCAAGAGTGGCCGTGTGTGAACAAGTGTGTT	1839
OY	841	CCGTGTGTGGCCGGGGCATGGAACCGAGGTGGCTTTGTGTGAGAGAAACAGAGTGGCCCTGG	900
Db	1840	CCGTGTGTGGCCGGGGCATGGAACCGAGGTGGCTTTGTGTGAGAGAAACAGAGTGGCCCTGG	1839
OY	901	GGGGCCCCCTTGACCTTTGATGAAACAGATGGAGAGAAAGAGAGAGACGTGAGAGACC	960
Db	1900	GGGGCCCCCTTGACCTTTGATGAAACAGATGGAGAGAAAGAGAGAGACGTGAGAGACC	1959
OY	961	TGGAAAAATGGATTAAAGCGAGACACGTCAGGAGGCGAGATCGGCCCTTAC	1012
Db	1960	TGGAAAAATGGATTAAAGCGAGACACGTCAGGAGGCGAGATCGGCCCTTAC	2011

## RESULT 8

ID	ABK64822 standard; DNA; 2390 BP.
XX	
AC	ABK64822;
XX	
DT	18-JUN-2002 (first entry)
XX	
DE	Human benign prostatic hyperplasia gene #717.
XX	
KM	Human; benign prostatic hyperplasia; BPH; prostate cancer; gene; ds.
XX	
OS	Homo sapiens.
XX	
PN	WO200212440-A2.
XX	
PD	14-FEB-2002.
XX	
PF	07-AUG-2001; 2001WO-US024708.
XX	
PR	07-AUG-2000; 2000US-0223333P.
PR	05-JUN-2001; 2001US-00873319.
XX	
XX	(GENE-) GENE LOGIC INC.
PA	(NISB ) JAPAN TOBACCO INC.
XX	
PI	Munger WE, Kulikarni P, Getzenberg RH, Waga I, Yamamoto J;
XX	
DR	WPI; 2002-257476/30.
XX	
PT	Identifying drugs for and diagnosing benign prostatic hyperplasia, by
PT	detecting expression levels of one or more genes in prostate cells from
PT	patient that are differentially regulated compared to normal prostate
PT	cells.
XX	
PS	Disclosure; Page 399-400; 444pp; English.
XX	
CC	The invention relates to a method of diagnosing (I) the onset or
CC	progression of benign prostatic hyperplasia (BPH), or screening (II) for

CC or identifying an agent that modulates the onset or progression of BPH.  
CC The method is based on changes in gene expression in BPH tissue isolated  
CC from patients exhibiting different clinical states of prostate  
CC hyperplasia as compared to normal prostate tissue. (I) comprises  
CC detecting the expression levels of one or more genes in prostate cells  
CC from the subject that are differentially regulated compared to normal  
CC prostate cells. (II) comprises preparing a first gene expression profile  
CC of BPH cells or BPH-like cell population, exposing the cells to the  
CC agent, preparing a second gene expression profile of the agent exposed  
CC cells, and comparing the first and second gene expression profiles. (I)  
CC is useful for diagnosing the onset or progression of BPH. (II) is useful  
CC for identifying an agent that modulates the onset or progression of BPH.  
CC The methods are useful to present information identifying the expression  
CC level in a tissue or cells, by comparing the expression level of genes  
CC given in the specification in the tissue or cells to the level of  
CC expression of gene in the database, and displaying the expression levels  
CC of at least one gene in the tissue or cell sample compared to the  
CC expression level in BPH. Agents using (II) are useful for treating BPH or  
CC prostate cancer. ABR654106-ABR64860 represent human benign prostatic  
CC hyperplasia gene sequences of the invention

XX  
SQ Sequence 2390 BP; 645 A; 589 C; 668 G; 488 T; 0 U; 0 Other;

Query Match 99.8%; Score 1010.4; DB 6; Length 2390;  
Best Local Similarity 99.8%; Pred. No. 7.5e-233;  
Matches 1011; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY	1	GGAGACGGCGTGGAAATGGGGACCGGGAGACGAGCGTGGCTTCAATAAAGCTGGTGGAC	60
Db	1000	AGGACGGCGGTGGAAATGGGACCGCGTGGAGACGAGGTGGCTTCAATAAAGCTGGTGGAC	1059
QY	61	CATGATGTAAGGACCAAGATCTTGATCTAGGCCCACTGTAGATCCAGATGAAAGCTCTGA	120
Db	1060	CATGATGTAAGGACCAAGATCTTGATCTTAGGCCCTCTGTAGATCCAGATGAAAGCTCTGA	1119
QY	121	CAACAGTGCATTTATGTACAAGATTAAATGACAGGTGTGACTTAAATGATCTGGCAGAA	180
Db	1120	CAACAGTGCATTTATGTACAAGATTAAATGACAGGTGTGACTTAAATGATCTGGCAGAA	1179
QY	181	CTTCTTTAAGCAGGTGGGGGTGTTAAGATGAACAAGAACTGGGCAACCATGATATCCA	240
Db	1180	CTTCTTTAAGCAGGTGGGGGTGTTAAGATGAACAAGAACTGGGCAACCATGATATCCA	1239
QY	241	CATCTACTGTGACAAGGAAACAGGAAAGCCCAAGGCGATGCCACAGTGTCTTATGANA	300
Db	1240	CATCTACTGTGACAAGGAAACAGGAAAGCCCAAGGCGATGCCACAGTGTCTTATGANA	1299
QY	301	CCCAACCACTGCCAAGGCTGCCGTGGAAATGGTTGAATGGGAAACATTTTCAAGGAGAA	360
Db	1300	CCCAACCACTGCCAAGGCTGCCGTGGAAATGGTTGAATGGGAAACATTTTCAAGGAGAA	1359
QY	361	ACTTAAAGTCTCCCTTGCTGGAGAAAGCCCTCAATGAAACGTAATGGCGGGTGTCTGCC	420
Db	1360	ACTTAAAGTCTCCCTTGCTGGAGAAAGCCCTCAATGAAACGTAATGGCGGGTGTCTGCC	1419
QY	421	ACCCCGTAGGGGACAGGACATGCCACCACTCCGTGGAGTCCAGAGGSCCAAGAGG	480
Db	1420	ACCCCGTAGGGGACAGGACATGCCACCACTCCGTGGAGTCCAGAGGSCCAAGAGG	1479
QY	481	TCCTGGGGGAAACCATGGGTGGCATGGGAGGCGGTGGAGAGATAGAAGAGGCTTCCCTCC	540
Db	1480	TCCTGGGGGAAACCATGGGTGGCATGGGAGGCGGTGGAGAGATAGAAGAGGCTTCCCTCC	1539
QY	541	AAGAGGACCCCGGGGGTTCGAGAGGAAACCCCTCTGAGAGAGAAACGTCCAGACCCGAGC	600
Db	1540	AAGAGGACCCCGGGGGTTCGAGAGGAAACCCCTCTGAGAGAGAAACGTCCAGACCCGAGC	1599
QY	601	TGAGACTGGCAGTGTCCCAATCCGGGTGTGTGAACAAGAACTTCCGCTGAGAAACAGA	660
Db	1600	TGAGACTGGCAGTGTCCCAATCCGGGTGTGTGAACAAGAACTTCCGCTGAGAAACAGA	1659
QY	661	GTGCAACCAAGTGTAAAGCCCAAAAGCTGAAAGCTTCTCCCGCACCCCTTTCGGCCCCC	720



Db 1960 TGAATAATGATTAAGGCGAGCACCCTGAGAGCCGAGATCGCCCTTAC 2011

RESULT 10  
ACNA0903  
ID ACNA0903 standard; cDNA; 2390 BP.

AC ACNA0903;

DT 18-NOV-2004 (first entry)

DE Tumour-associated antigenic target (TAT) cDNA DNA269830, SEQ ID NO:5989.

XX Tumour-associated antigenic target; TAT; human; overexpression; cancer;  
XX tumour; diagnosis; cell proliferative disorder; breast cancer;  
XX colorectal cancer; lung cancer; ovarian cancer; liver cancer;  
XX central nervous system cancer; bladder cancer; pancreatic cancer;  
XX cervical cancer; melanoma; leukaemia; hybridisation probe;  
XX chromosome identification; chromosome mapping; gene mapping;  
XX gene therapy; cytostatic; gene; ss.

OS Homo sapiens.

XX WO2004030615-A2.

PN 15-APR-2004.

XX 29-SEP-2003; 2003WO-US028547.

XX 02-OCT-2002; 2002US-0414971P.

XX (GENTH) GENENTECH INC.

XX Wu TD, Zhang Z, Zhou Y;

XX WPI; 2004-347921/32.

XX P-PSDB; ABM62330.

PT New tumor-associated antigenic target polypeptides and nucleic acids,  
PT useful in preparing a medicament for treating or detecting a  
PT proliferative disorder, e.g. breast, lung, colorectal, ovarian or  
PT prostate cancer or tumor.

XX Claim 1; SEQ ID NO 5989; 7273bp; English.

XX The invention relates to human tumour-associated antigenic target (TAT)  
XX polypeptides, and their related nucleic acids. The TAT polypeptides are  
XX overexpressed in cancer tissues compared to normal tissues, and may thus  
XX serve as effective targets for the diagnosis and treatment of cancer in  
XX mammals. The invention also relates to the TAT nucleic acids and  
XX sequences at least 80% identical to the TAT nucleic acids and  
XX polypeptides; expression vectors and host cells comprising a TAT nucleic  
XX acid; an antibody specific for a TAT polypeptide; a peptide or organic  
XX molecule which binds to a TAT polypeptide; fusion proteins comprising a  
XX TAT polypeptide; and methods and compositions for the treatment or  
XX diagnosis of cancer in mammals. TAT polypeptides, nucleic acids,  
XX antibodies, antigens, binding molecules and compositions are useful  
XX for diagnosing or treating a cell proliferative disorder associated with  
XX increased TAT expression, particularly cancers such as breast cancer,  
XX colorectal cancer, lung cancer, ovarian cancer, liver cancer, bladder  
XX cancer, pancreatic cancer, cervical cancer, cancers of the central  
XX nervous system, melanoma and leukaemia. TAT nucleic acids may further be  
XX used as hybridisation probes, in chromosome and gene mapping, in  
XX chromosome identification and in gene therapy. The present sequence  
XX represents a TAT nucleic acid of the invention.

SO Sequence 2390 BP; 645 A; 589 C; 668 G; 488 T; 0 U; 0 Other;

Query Match 99.8%; Score 1010.4; DB 13; Length 2390;  
Best Local Similarity 99.9%; Pred. No. 7.5e-233;  
Matches 1011; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AGAAGCGGTGGAATGAGCAGGCTGAGAGGAGGTGCTTCAATAGCCTGTGAGC 60  
Db 1000 AGAGACCGGTGGAATGAGCAGGCTGAGAGGAGGTGCTTCAATAGCCTGTGAGC 1059  
QY 61 CATGATGAAAGAACCAAGATCTTGATCTAGAGCCCACTGTAGATCCAGATGAGACTGTGA 120  
Db 1060 CATGATGAAAGAACCAAGATCTTGATCTAGAGCCCTCTGTAGATCCAGATGAGACTGTGA 1119  
QY 121 CAACGTGCAATTTATGTCAGAGATTAATGACAGTGTGACTCTGATGATCTGACAGA 180  
Db 1120 CAACGTGCAATTTATGTCAGAGATTAATGACAGTGTGACTCTGATGATCTGACAGA 1179  
QY 181 CTCTCTTAAGCAGTGTGAGGCTGTTAAGATGAAACAAGAACTGGCAACCCATGATCCA 240  
Db 1180 CTCTCTTAAGCAGTGTGAGGCTGTTAAGATGAAACAAGAACTGGCAACCCATGATCCA 1239  
QY 241 CATCTACTGGAACAGAGAAACAGAAAGCCCAAGGCGATGCCAGTGTCTTATGAGA 300  
Db 1240 CATCTACTGGAACAGAGAAACAGAAAGCCCAAGGCGATGCCAGTGTCTTATGAGA 1299  
QY 301 CCCACCACTGGCCAGAGCGCTGGATGATGTTGATGAGGAAAGATTTCAAGGAGCA 360  
Db 1300 CCCACCACTGGCCAGAGCGCTGGATGATGTTGATGAGGAAAGATTTCAAGGAGCA 1359  
QY 361 ACTTAAGTCTCCCTTCTCTGAGAGAGCCTCCATGAAACAGTATGCGGGTGTCTGCC 420  
Db 1360 ACTTAAGTCTCCCTTCTCTGAGAGAGCCTCCATGAAACAGTATGCGGGTGTCTGCC 1419  
QY 421 ACCCGGTGAGGAGGAGGAGCAGTATGACCACTCCCTGTGAGAGTCTCAGAGGCCCCAGAGG 480  
Db 1420 ACCCGGTGAGGAGGAGGAGCAGTATGACCACTCCCTGTGAGAGTCTCAGAGGCCCCAGAGG 1479  
QY 481 TCCGTGGGGAGCCCATGAGGTGCAATGAGAGCGCTGAGAGGATGAGAGGCTTCCCTCC 540  
Db 1480 TCCGTGGGGAGCCCATGAGGTGCAATGAGAGCGCTGAGAGGATGAGAGGCTTCCCTCC 1539  
QY 541 AAGAGAGCCCGGGGCTTCCGAGAGAACCCCTCTGAGAGGAGAAAGTCCAGACACGAGC 600  
Db 1540 AAGAGAGCCCGGGGCTTCCGAGAGAACCCCTCTGAGAGGAGAAAGTCCAGACACGAGC 1599  
QY 601 TGGAGACTGAGCAGTGTCCCATATCCGGGTTGTGGAACCAAGAACTTGCTGAGAGACAGA 660  
Db 1600 TGGAGACTGAGCAGTGTCCCATATCCGGGTTGTGGAACCAAGAACTTGCTGAGAGACAGA 1659  
QY 661 GTGCAACAGTGTAAAGCCCAAGGCTTCCCTGAGAGGCTTCCCTGAGAGGCTTCCCTGAGAGG 720  
Db 1660 GTGCAACAGTGTAAAGCCCAAGGCTTCCCTGAGAGGCTTCCCTGAGAGGCTTCCCTGAGAGG 1719  
QY 721 GGGTGTGATCTGAGAGAGTGTGCTGTGATGAGAGGAGAGAGAGTGTGCTCAT 780  
Db 1720 GGGTGTGATCTGAGAGAGTGTGCTGTGATGAGAGGAGAGAGTGTGCTCAT 1779  
QY 781 GATCTGTGTGTGCTGCGGTGAAATGTTCAAGAGTGTGCTGTGAGAGAGAGTGTGCTT 840  
Db 1780 GATCTGTGTGTGCTGCGGTGAAATGTTCAAGAGTGTGCTGTGAGAGAGAGTGTGCTT 1839  
QY 841 CCGTGTGTGCGCGGGGAGATGAGACGAGTGTGCTTGTGAGAGAGAGAGTGTGCTT 900  
Db 1840 CCGTGTGTGCGCGGGGAGATGAGACGAGTGTGCTTGTGAGAGAGAGAGTGTGCTT 1899  
QY 901 GGGGCCCCCTGAGCTTGTGATGAGAGATGAGAGAGAGAGAGAGAGTGTGAGAGC 960  
Db 1900 GGGGCCCCCTGAGCTTGTGATGAGAGATGAGAGAGAGAGAGAGAGTGTGAGAGC 1959  
QY 961 TGAATAATGATTAAGGCGAGCAGCTCAGAGGCGCAGAGATCGCCCTTAC 1012  
Db 1960 TGAATAATGATTAAGGCGAGCAGCTCAGAGGCGCAGAGATCGCCCTTAC 2011

RESULT 11  
ADRO7446  
ID ADRO7446 standard; cDNA; 2026 BP.



AC ADR07446;  
 XX  
 DT 04-NOV-2004 (first entry)  
 XX  
 DE Full length human cDNA useful for treating neurological disease Seq 952.  
 XX  
 KM Gene; ssr; human; oligo-capping method; diagnostic marker; gene therapy;  
 KM osteoporosis; neurological disease; Alzheimer's disease;  
 KM Parkinson's disease; dementia; short memory; cancer;  
 KM sense or motor function; emotional reaction; fear response; panic;  
 KM osteopathic; neuroprotective; nootropic; antiparkinsonian; cyostatic;  
 KM tranquilliser.  
 XX  
 OS Homo sapiens.  
 XX  
 PN BP1447413-A2.  
 XX  
 PD 18-AUG-2004.  
 XX  
 PP 12-FEB-2004; 2004EP-00003145.  
 XX  
 PR 14-FEB-2003; 2003JP-00102207.  
 PR 09-MAY-2003; 2003JP-00131452.  
 XX  
 PA (REAS-) RES ASSOC BIOTECHNOLOGY.  
 XX  
 PI Isogai T, Yamamoto J, Nishikawa T, Isono Y, Sugiyama T, Otsuki T,  
 PI Wakamatsu A, Ishii S, Nagai K, Irie R;  
 XX  
 DR WPI; 2004-583265/57.  
 DR P-PSDB; ADR09402.  
 XX  
 PT New 1995 cDNA, useful for treating osteoporosis, neurological diseases,  
 PT Alzheimer's diseases, Parkinson's diseases, dementia and various cancers.  
 XX  
 PS Claim 1; SEQ ID NO 952; 2686bp; English.  
 XX  
 CC This invention relates to novel, isolated full length human cDNA  
 CC molecules and the encoded proteins thereof. Specifically, it refers to  
 CC cDNA clones obtained by an oligo-capping method, where none of these  
 CC clones are identical to any known human mRNAs. The present invention  
 CC describes an immunosay to identify agonists and antagonists, as well as  
 CC antibodies, antisense molecules and siRNAs that can all be used to bind  
 CC to and modulate expression of the cDNA molecules. As such, these  
 CC molecules are useful for diagnostic markers or therapeutic targets for  
 CC the various diseases or morbid states. In particular, they are useful in  
 CC gene therapy for treating osteoporosis, neurological disease, Alzheimer's  
 CC disease, Parkinson's disease, dementia, short memory and various cancers,  
 CC as well as for maintaining equilibrium of sense or motor function, and  
 CC for treating emotional reaction, fear response and panic. Accordingly,  
 CC they exhibit osteopathic, neuroprotective, nootropic, antiparkinsonian,  
 CC cyostatic and tranquilliser activities. This polynucleotide is a full  
 CC length human cDNA sequence of the invention. NOTE: This sequence is not  
 CC given in the sequence listing of the specification but can be obtained on  
 CC CD-ROM from the European Patent Office, Vienna Sub-office.  
 XX  
 XX  
 SQ Sequence 2026 BP; 540 A; 513 C; 590 G; 383 T; 0 U; 0 Other;

Query Match 99.7%; Score 1008.8; DB 13; Length 2026;  
 Best Local Similarity 99.8%; Pred. No. 1,7e-232;  
 Matches 1010; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AGAAGCGGCTGGAGTGGGCAAGCGCTGGAGAGCGAGGTGGCTTCATTAAGCCTGGTGACC 60  
 DB 838 AGAAGCGGCTGGAGTGGGCAAGCGCTGGAGAGCGAGGTGGCTTCATTAAGCCTGGTGACC 897  
 QY 61 CATGATGAAGGACCAATCTTGATCTAAGGCCCACTGGTGAATCCCAATGAATGAAGCTCTGA 120  
 DB 898 CATGATGAAGGACCAATCTTGATCTAAGGCCCACTGGTGAATCCCAATGAATGAAGCTCTGA 957  
 QY 121 CAACAGTCAATTTATATACAGATTAATGAACAGTGTGACTCTAGATGATCTGACAGA 180  
 DB 958 CAACAGTCAATTTATATACAGATTAATGAACAGTGTGACTCTAGATGATCTGACAGA 1017

QY 181 CTTCTTTAAGCACTGTGGGTGTGTTAAGATGAACAAGAGAACTGGGCAACCATGATCCA 240  
 DB 1018 CTTCTTTAAGCACTGTGGGTGTGTTAAGATGAACAAGAGAACTGGGCAACCATGATCCA 1077  
 QY 241 CATCTACCTGGACAGAAACAGAAAGCCCAAGGCGATGCCACAGTGTCTCTATGAGA 300  
 DB 1078 CATCTACCTGGACAGAAACAGAAAGCCCAAGGCGATGCCACAGTGTCTCTATGAGA 1137  
 QY 301 CCCAACCACTGCCAAGGCTGCGTGAATGTTGATGAGAAAGATTTCAAGGACAA 360  
 DB 1138 CCCAACCACTGCCAAGGCTGCGTGAATGTTGATGAGAAAGATTTCAAGGACAA 1197  
 QY 361 ACTTTAAGTCTCCCTTGTCTGTGGAAGAAAGCTTCCATTAATGAATGAGGAGTGTGCC 420  
 DB 1198 ACTTTAAGTCTCCCTTGTCTGTGGAAGAAAGCTTCCATTAATGAATGAGGAGTGTGCC 1257  
 QY 421 ACCCGGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 480  
 DB 1258 ACCCGGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1317  
 QY 481 TCTTGGGGAACCATGAGGTGCAATGAGAGGCGGTGAGAGATGAGAGAGCTTCCCTCC 540  
 DB 1318 TCTTGGGGAACCATGAGGTGCAATGAGAGGCGGTGAGAGATGAGAGAGCTTCCCTCC 1377  
 QY 541 AAGAGACCCCGGGGTTCCCGAGGAAACCCCTCTGAGAGAGAAAGCTCCAGACCGAGAC 600  
 DB 1378 AAGAGACCCCGGGGTTCCCGAGGAAACCCCTCTGAGAGAGAAAGCTCCAGACCGAGAC 1437  
 QY 601 TGGAGACTGGCAGTGTCCCAATCCGGGTGTGGAACACAGAACTTGCTGAGAAACAGA 660  
 DB 1438 TGGAGACTGGCAGTGTCCCAATCCGGGTGTGGAACACAGAACTTGCTGAGAAACAGA 1497  
 QY 661 GTGCAACCACTGTAAAGCCCAAGCCTGAAGCTTCTCCGCAACCTTTCGACCCC 720  
 DB 1498 GTGCAACCACTGTAAAGCCCAAGCCTGAAGCTTCTCCGCAACCTTTCGACCCC 1557  
 QY 721 GGGTGTGATGTCGTGAGAGAGTGGCTGTGTCATTCGGGAGAGAAAGAGTGGCTCAT 780  
 DB 1558 GGGTGTGATGTCGTGAGAGAGTGGCTGTGTCATTCGGGAGAGAAAGAGTGGCTCAT 1617  
 QY 781 GAATCGTGTGTCCTCCGAGTGAATGTTCAAGAGTGGCGGTGAGAGACAGAGTGGCTT 840  
 DB 1618 GAATCGTGTGTCCTCCGAGTGAATGTTCAAGAGTGGCGGTGAGAGACAGAGTGGCTT 1677  
 QY 841 CCGTGTGAGCGCGGAGCATGAGACCGAGTGGCTTGTGAGAGAAAGAGAGTGGCTTGG 900  
 DB 1678 CCGTGTGAGCGCGGAGCATGAGACCGAGTGGCTTGTGAGAGAAAGAGAGTGGCTTGG 1737  
 QY 901 GGGGCCCTCTGACCTTTGATGGAACAGATGGAGAGAAAGAGAGAGAGAGAGAGAGAG 960  
 DB 1738 GGGGCCCTCTGACCTTTGATGGAACAGATGGAGAGAAAGAGAGAGAGAGAGAGAGAG 1797  
 QY 961 TGGAAAAATGATTAAGGAGAGACCGTCAAGAGCGGAGAGATCGGCGCTTAC 1012  
 DB 1798 TGGAAAAATGATTAAGGAGAGACCGTCAAGAGCGGAGAGATCGGCGCTTAC 1849

RESULT 12  
 ADP56333  
 ID ADP56333 standard, cDNA; 1807 BP.

XX ADP56333;  
 XX  
 DT 18-NOV-2004 (first entry)  
 XX

DE Human PRO cDNA sequence SEQ ID NO:2309.

XX human, PRO; immune related disease; inflammatory immune response;  
 XX immune response stimulative; antiallergic; antianaemic; antidiarrhetic;  
 KM antiseptic; antidiabetic; antiinflammatory; antiparasitic;  
 KM antipneumatic; antichyroid; CNS; dermatological; gastrointestinal;  
 KM haemostatic; hepatotropic; immunostimulant; immunosuppressive; muscular;



XX MO2004030615-A2.  
 PN 15-APR-2004.  
 PD 29-SEP-2003; 2003MO-US028547.  
 XX 02-OCT-2002; 2002US-0414971P.  
 PR (GENTH ) GENENTECH INC.  
 PA Wu TD, Zhang Z, Zhou Y;  
 PI WPI: 2004-347921/32.  
 DR P-PSDB; ABM60157.  
 XX New tumor-associated antigenic target polypeptides and nucleic acids,  
 PT useful in preparing a medicament for treating or detecting a  
 PT proliferative disorder, e.g. breast, lung, colorectal, ovarian or  
 PT prostate cancer or tumor.  
 XX Claim 1; SEQ ID NO 490; 7273bp; English.  
 PS The invention relates to human tumour-associated antigenic target (TAT)  
 CC polypeptides, and their related nucleic acids. The TAT polypeptides are  
 CC overexpressed in cancer tissues compared to normal tissues, and may thus  
 CC serve as effective targets for the diagnosis and treatment of cancer in  
 CC mammals. The invention also relates to nucleic acid and polypeptide  
 CC sequences at least 80% identical to the TAT nucleic acids and  
 CC polypeptides; expression vectors and host cells comprising a TAT nucleic  
 CC acid; an antibody specific for a TAT polypeptide; a peptide or organic  
 CC molecule which binds to a TAT polypeptide; fusion proteins comprising a  
 CC TAT polypeptide; and methods and compositions for the treatment or  
 CC diagnosis of cancer in mammals. TAT polypeptides, nucleic acids,  
 CC antibodies, antagonists, binding molecules and compositions are useful  
 CC for diagnosing or treating a cell proliferative disorder associated with  
 CC increased TAT expression, particularly cancers such as breast cancer,  
 CC colorectal cancer, lung cancer, ovarian cancer, liver cancer, bladder  
 CC cancer, pancreatic cancer, cervical cancer, cancers of the central  
 CC nervous system, melanoma and leukaemia. TAT nucleic acids may further be  
 CC used as hybridisation probes, in chromosome and gene mapping, in  
 CC chromosome identification and in gene therapy. The present sequence  
 CC represents a TAT nucleic acid of the invention  
 XX  
 XX Sequence 2177 BP; 602 A; 539 C; 605 G; 431 T; 0 U; 0 Other;  
 SQ  
 Query Match 89.7%; Score 908; DB 13; Length 2177;  
 Best Local Similarity 94.9%; Pred. No. 3,1e-208;  
 Matches 960; Conservative 0; Mismatches 50; Indels 2; Gaps 2;

QY 361 ACTTAAGTCTCCCTTCTCGGAAGAAGCCTCCATGAACAGTATGCGGGGTGCTGCC 420  
 DB 1350 ACTTAAGTCTCTCTTCTCGGAAGAGCCTCCAGTGAACAGTATGAGGGTGTATGCC 1409  
 QY 421 ACCCGTGAAGGAGAGAGAGTGCACCACTCCGTGAGTCCAGAGAGGCCAGAGG 480  
 DB 1410 ACCCCATGAGGAGAGAGAGTGCACCACTCTGCGAGGTCCAGAGAGGCCAGAG 1469  
 QY 481 TCCTGAGGAGCCCATGAGGTGCATGAGAGGCCGTGAGAGAGATGAGAGGCTTCCCTCC 540  
 DB 1470 TCCTGAGGAGCCCATGAGGTGCATGAGAGGCCGTGAGAGAGATGAGAGGCTTCCCTCC 1529  
 QY 541 AAGAGACCCCGGGGTTCCCGAGAGAACCCCTCTGAGAGAGAGAAAGTCCAGACCCAGC 600  
 DB 1530 AAGAGACCCCGAGGGTTCCCGAGAGAACCTCTGAGAGAGAGAAAGTCCAGACCCAGC 1589  
 QY 601 TGGAGACTGACAGTGTCCCAATCCGGGTTGAGAAACAGAACTTGCTGAGAGACAGA 660  
 DB 1590 TGGAGACAGGCAATGTCCCAATCCGGGTTGAGAAACAGAACTTGCTGAGAGACAGA 1649  
 QY 661 GTGCAACCAAGTGTAAAGCCCAAGCCTGAAGGCTTCTCCGCAACCTTTCCGCCCC 720  
 DB 1650 GAGCAACAGTGTAAAGCTCCAAAGCCTGAAGGCTTCTCCGCAACCTTTCCGCCCC 1709  
 QY 721 GGGTGTGATCTGTGAGAGAGTGGCCTGTGTGCAATGCGGGAGAGAAAGAGTGGCTCAT 780  
 DB 1710 GGGTGTGATCAATGAGAGAGTGGCCTGTGTGCAATGCGGGAGAGAAAGAGTGGCTCAT 1769  
 QY 781 GGATCGTGTGATCCCGGTGGAATGTTCAAGAGTGGCCGAGTGTGAGACAGAGGTGGCTT 840  
 DB 1770 GGATCGTGTGATCCCGGTGGAATGTTCAAGAGTGGCTGTGTGTGAGACAGAGGTGGCTT 1829  
 QY 841 CCGTGTGAGCCGAGGAGATGAGACCGAGGTGCTTTGTGTGAGAGAAAGAGTGGCTGG 900  
 DB 1830 CTGTGTGTGAG-CTGAGGATGAGACCGAGGTGCTTTGTGTGAGAGAAAGAGTGGCTGG 1888  
 QY 901 GGGGCCCCCTGAGCTTTGATGAGACAGATGAGAGAGAGAGAGAGAGTGGAGAGC 960  
 DB 1889 GGGGCCCCCT-GGACTTTGATGATGAGACAGATGAGAGAGAGAGAGAGAGTGGAGAGC 1947  
 QY 961 TGGAAAAATGATTAAGAGGAGAGACCGTCAAGAGCGAGAGATCGGCCCTTAC 1012  
 DB 1948 TGGAAAAATGATTAAGAGGAGAGACCTGTCAAGAGCGAGAGATCGGCCCTTAC 1999

RESULT 14  
 AAS62623/C  
 ID AAS62623 standard; cDNA; 2273 BP.  
 XX AAS62623;  
 AC  
 XX  
 DT 14-FEB-2002 (first entry)  
 XX  
 DE cDNA sequence #410 encoding novel human secreted protein.  
 XX  
 XX Human secreted protein; hyperproliferative disorder; autoimmune disorder;  
 XX immune deficiency disorder; blood disorder; inflammatory disorder;  
 XX infectious disorder; gene therapy; antimicrobial; hepatotropic;  
 XX immunosuppressive; antirheumatic; ss.  
 XX  
 OS Homo sapiens.  
 OS  
 PN WO200177291-A2.  
 PN  
 PD 18-OCT-2001.  
 PD  
 XX 29-MAR-2001; 2001MO-US010485.  
 PF  
 XX 06-APR-2000; 2000US-0195604P.  
 PR  
 XX  
 PA (GENY ) GENETICS INST INC.  
 XX

PI Wong GG, Clark HF, Fechtel K, Agostino MJ, Howes SH, Resnick RJ;  
 PI Gulkota K, Graham UR;  
 DR WPI, 2002-010900/01.  
 PT New polynucleotides encoding secreted proteins useful for treating e.g.  
 PT asthma, HIV and Crohn's disease.  
 XX  
 PS Claim 1, Page 291, 391pp; English.  
 CC The present invention relates to the isolation of novel cDNA sequences  
 CC which encode human secreted proteins. The cDNA sequences have been  
 CC derived from a variety of human tissues. The invention also provides a  
 CC method for producing proteins from these polynucleotide sequences. The  
 CC proteins are useful for identifying compounds that modulate their  
 CC activity and production, and the cell is also useful for identifying  
 CC compounds that modulate expression of the polynucleotide sequences  
 CC encoding the secreted proteins. The sequences of the invention are useful  
 CC for treating diseases such as hyperproliferative disorders (e.g. cancer),  
 CC immune deficiency disorders (e.g. severe combined immunodeficiency  
 CC (SCID)), autoimmune disorders (e.g. multiple sclerosis), blood disorders  
 CC (e.g. thrombocytopenia), inflammatory disorders (e.g. arthritis) and  
 CC infectious disorders (e.g. hepatitis). The polynucleotide sequences of  
 CC the invention are also useful in gene therapy. AA62214-AA62838  
 CC represent the cDNA sequences of the invention that encode for novel human  
 CC secreted proteins  
 XX  
 SQ Sequence 2273 BP; 477 A; 636 C; 562 G; 598 T; 0 U; 0 Other;  
 Query Match 87.2%; Score 882.4; DB 6; Length 2273;  
 Best Local Similarity 94.6%; Pred. No. 4.6e-202;  
 Matches 957; Conservative 0; Mismatches 1; Indels 54; Gaps 2;  
 QY 1 AGAAGCGGCTGGATGCGCAGCGCTGGAGAGCGAGTGGCTTCAATAGCTGTGAGCC 60  
 DB 1310 AGAAGCGGCTGGATGCGCAGCGCTGGAGAGCGAGTGGCTTCAATAGCTGTGAGCC 1254  
 QY 61 CATGATGAGAGAGCGATCTTGATCTAGGCGCCAGCTGATCCAGATGAAGACTGTA 120  
 DB 1253 CATGATGAGAGAGCGATCTTGATCTAGGCGCCAGCTGATCCAGATGAAGACTGTA 1194  
 QY 121 CAACAGTCAATTATGATCAAGATTTAAATGACAGTGTACTCTAGATGCTGGAGA 180  
 DB 1193 CAACAGTCAATTATGATCAAGATTTAAATGACAGTGTACTCTAGATGCTGGAGA 1134  
 QY 181 CTCTTTAAGCAAGTGGGGTGTAAATGACAGAGAACTGGGCAACCATGATCCA 240  
 DB 1133 CTCTTTAAGCAAGTGGGGTGTAAATGACAGAGAACTGGGCAACCATGATCCA 1074  
 QY 241 CATCTACCTGAGAAAGAGAAAGAGAAAGCCAAAGGAGTCCAGTGTCTATGAGA 300  
 DB 1073 CATCTACCTGAGAAAGAGAAAGAGAAAGCCAAAGGAGTCCAGTGTCTATGAGA 1014  
 QY 301 CCCAGCCACTGCAAGGCTCCGTGATGATGAGAAAGATTTCAAGGAGCAA 360  
 DB 1013 CCCAGCCACTGCAAGGCTCCGTGATGATGAGAAAGATTTCAAGGAGCAA 954  
 QY 361 ACTTAAAGTCTCCCTTCTCTGGAAGAAAGCTTCAATGAAGTATGCGGGGTGCTGCC 420  
 DB 953 ACTTAAAGTCTCCCTTCTCTGGAAGAAAGCTTCAATGAAGTATGCGGGGTGCTGCC 894  
 QY 421 ACCCGGAGAGGAG 480  
 DB 893 ACCCGGAGAGGAG 834  
 QY 481 TCTGAGGAG 540  
 DB 833 TCTGAGGAG 774  
 QY 541 AAG 600  
 DB 773 AAG 714

QY 601 TGGAGACTGAG 660  
 DB 713 TGGAGACTGAG 654  
 QY 661 GTGCAACAG 720  
 DB 653 GTGCAACAG 643  
 QY 721 GGGTGG 780  
 DB 642 --GTGATGATCGTGGAG 585  
 QY 781 GATCGTGG 840  
 DB 584 GATCGTGG 525  
 QY 841 CCGTGG 900  
 DB 524 CCGTGG 465  
 QY 901 GGGGCCCCCTGACCTTGTATGAG 960  
 DB 464 GGGGCCCCCTGACCTTGTATGAG 405  
 QY 961 TGGAAATATGATTAAG 1012  
 DB 404 TGGAAATATGATTAAG 353  
 RESULT 15  
 AB199383  
 ID AB199383 standard; cDNA; 2188 BP.  
 XX  
 AC AB199383;  
 XX  
 DT 07-MAR-2002 (first entry)  
 XX  
 DE Mouse ischaemic condition related cDNA sequence SEQ ID NO:288.  
 XX  
 KW Mouse; ischemia; compressive ischemia; occlusive ischemia;  
 KW vasospastic ischemia; ischaemic condition; ischaemic disease; ss.  
 XX  
 OS Mus musculus.  
 XX  
 PN M0200188188-A2.  
 XX  
 PD 22-NOV-2001.  
 XX  
 PF 18-MAY-2001; 2001WO-JP004192.  
 XX  
 PR 18-MAY-2000; 2000JP-00145977.  
 XX  
 PA (UNINT-) UNIV NITHON SCHOOL JURIDICAL PERSON.  
 XX  
 PI Iehikawa K, Arai S, Takahashi Y, Nagata T, Ishii Y;  
 XX  
 DR WPI, 2002-034733/04.  
 PT P-PSDB; ABB57126.  
 XX  
 PT Examining the ischemic condition (e.g. occlusive ischemia) by measuring  
 PT expression levels of particular genes defined in the specification or by  
 PT determining the expression profile of a gene group comprising these  
 PT genes.  
 XX  
 PS Claim 2, Page 794-799; 2690pp; English.  
 XX  
 CC The present invention describes a method for examining ischaemic  
 CC conditions, comprising measuring the expression levels of particular  
 CC genes (I) in a test sample or determining the expression profile of a  
 CC gene group in the sample comprising genes rejected from (I). The method  
 CC is useful for examining the ischaemic condition (e.g. compressive  
 CC ischemia, occlusive ischemia or vasospastic ischemia) by measuring the  
 CC expression levels of particular genes (AB199202 to AB199912, encoding the

CC protein sequences in ABB57020 to ABB57274) or by determining the  
CC expression profile of a gene group comprising these genes. The expression  
CC levels or expression profiles produced by these genes are used as an  
CC indicator when screening for ischemic condition-improving drugs or  
CC therapeutic for ischemic diseases. AB19913 and AB19914 represent PCR  
CC primers for a mouse ischemic condition related sequence, which are used  
CC in the exemplification of the present invention  
XX

Sequence 2188 BP; 587 A; 551 C; 638 G; 412 T; 0 U; 0 Other;

Query Match 86.9%; Score 879.2; DB 6; Length 2188;  
Best Local Similarity 91.7%; Pred. No. 2,7e-201;  
Matches 928; Conservative 0; Mismatches 84; Indels 0; Gaps 0;

```
QY 1 AGAAGCGGTGTAATGGGACGCTGGAGAGCGAGTGGCTTCAATAAAGCTTGAGACC 60
DB 1010 AGAAGGACGGGTGATGCTGGGCGCTGGAGAGCGAGTGGCTTCAATAAAGCTTGAGACC 1069
QY 61 CATGATGAAGAGCCAGATCTTGAATCAGGCCCACTGTAGATCCAGATGAAGACTTGA 120
DB 1070 CATGATGAAGAGCCAGATCTTGAATCAGGCCCTTCTATGATCCAGATGAAGACTTGA 1129
QY 121 CAACAGTGAATTTATGTACAAAGATTAAATGACAGTGAATCTAGATGATCTGACAGA 180
DB 1130 CAACAGTGAATTTATGTACAAAGATTAAATGACAGTGAATCTAGATGATCTGACAGA 1189
QY 181 CTTCTTTAAGCAAGTGTGGGTTTGTAAATGAACAAAGAACTGGGCAACCCATGATCA 240
DB 1190 CTTCTTTAAGCAAGTGTGGGTTTGTCAATGAACAAAGAACTGGGCAACCCATGATCA 1249
QY 241 CATCTACCTGAGCAAGAAACAGAAAGCCCAAGGCGATGCCAGATGCTTATGAGA 300
DB 1250 TATCTACTGATTAAGAGAACAGAAAGCCCTAAAGGGGACGCAAGTGTCTATGAGA 1309
QY 301 CCCAGCCACTGCGCAAGGCTGCGTGAATGATGATGATGATGATGATGATGATGATGAT 360
DB 1310 TCCACCAACTGCGCAAGGCTGCGTGAATGATGATGATGATGATGATGATGATGATGAT 1369
QY 361 ACTTAAAGTCTCCCTTGTCTGGAAGAAAGCTTCAATGAACATATGCGGGGTGCTGCC 420
DB 1370 ACTTAAAGTCTCTTGTCCGAAAGAAAGCTTCAATGAACATATGCGGGGTGCTGCC 1429
QY 421 ACCCGGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 480
DB 1430 ACCCTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1489
QY 481 TCCGTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 540
DB 1490 CCGTGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1549
QY 541 AAGAGGACCCGGGGTTCCGAGGGAAAGCTTCTGAGAGGAAAGCTTCCAGCACCGAGC 600
DB 1550 AAGAGGGGCCCCGAGGGCTCCAGAGGAAAGCTTCTGAGAGGAAAGCTTCCAGCACCGAGC 1609
QY 601 TGAAGACTGCAAGTGTCCCAATCCGGGTGTGGAACCAAGAACTTGCCTGAGAAACAGA 660
DB 1610 TGAAGACTGCAAGTGTCCCAATCCGGGTGTGGAACCAAGAACTTGCCTGAGAAACAGA 1669
QY 661 GTGCAACCAAGTAAAGGCCCAAGAGCTTGAAGGCTTCTCCGCGCACCTTTCCGAGCC 720
DB 1670 ATGCAACCAAGTAAAGGCCCTTAAGCCGAGGGCTTCTCCGCGCACCTTTCCGAGCC 1729
QY 721 GAGTGTGATCTGTGCAAGAGTGGCCCTGTGTGCAATGCGGAGAGAAAGAGTGGCTCAT 780
DB 1730 GAGTGTGATCTGTGCAAGAGTGGCCCTGTGTGCAATGCGGAGAGAAAGAGAGTGGCTCAT 1789
QY 781 GATCTGTGTGTGCTCCCGTGAATGTTCAGAGGTGCGCTGTGTGAGACAGAGTGGCTT 840
DB 1790 GATCTGTGTGTGCTCCCGTGAATGTTCAGAGGTGCGCTGTGTGAGACAGAGTGGCTT 1849
QY 841 CCGTGTGTGCGCGGAGCAATGACCGAGGTGCTTTGTGTGAGAAAGAGAGTGGCTTGG 900
DB 1850 CCGAGTGTGCGCTGTGAAATGACCGAGGTGCTTTGTGTGAGAAAGAGAGTGGCTTGG 1909
```

```
QY 901 GGGGCCCCCTTGAACCTTTGATGGAACAGATGGAGAGAAAGAGAGAGAGTGGAGACC 960
DB 1910 GGGGCCCCCTTGAACCTTTGATGGAACAGATGGAGAGAAAGAGAGAGAGTGGAGACC 1969
QY 961 TGAATAATGATTAAGGCGAGGACCGTCAAGAGCGGAGAGATCGGCTTAC 1012
DB 1970 TGAATAATGATTAAGGCGAGGACCGTCAAGAGCGGAGAGATCGGCTTAC 2021
```

Search completed: February 20, 2005, 05:30:50  
Job time : 495.802 secs

***This Page Blank (uspto)***



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 20, 2005, 04:13:53 ; Search time 167.828 Seconds  
(Without alignments)  
9866.749 Million cell updates/sec

Title: US-10-791-017a-1\_COPY\_1000\_2011

Perfect score: 1012  
Sequence: 1 aggcagcgatggaatggca.....agcgagagatcgccctac 1012

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0 %  
Maximum Match 100 %  
Listing first 45 summaries

Database : Issued Patents NA: \*  
1: /cgn2\_6/ptodata/1/ina/5A COMB.seq: \*  
2: /cgn2\_6/ptodata/1/ina/5B COMB.seq: \*  
3: /cgn2\_6/ptodata/1/ina/6A COMB.seq: \*  
4: /cgn2\_6/ptodata/1/ina/6B COMB.seq: \*  
5: /cgn2\_6/ptodata/1/ina/PTUS COMB.seq: \*  
6: /cgn2\_6/ptodata/1/ina/backfile1.seq: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1010.4	99.8	2371	2 US-08-343-443B-1	Sequence 1, Appl
2	956	94.5	1785	4 US-09-949-016-5043	Sequence 5043, Ap
3	956	94.5	1785	4 US-09-949-016-5044	Sequence 5044, Ap
4	911.2	90.0	6002	4 US-09-949-016-13696	Sequence 13696, A
5	860	85.0	1783	4 US-09-949-016-1954	Sequence 1954, Ap
6	338	33.4	411	4 US-09-621-976-13361	Sequence 13361, A
7	258.4	25.5	35784	4 US-09-949-016-16785	Sequence 16785, A
8	258.4	25.5	35784	4 US-09-949-016-16786	Sequence 16786, A
9	213	21.0	601	4 US-09-949-016-176641	Sequence 176641, A
10	213	21.0	601	4 US-09-949-016-17664	Sequence 17664, A
11	164.8	16.3	1939	4 US-09-919-039-322	Sequence 322, App
12	135.8	13.4	601	4 US-09-949-016-176637	Sequence 176637, A
13	135.8	13.4	601	4 US-09-949-016-176638	Sequence 176638, A
14	135.8	13.4	601	4 US-09-949-016-176660	Sequence 176660, A
15	135.8	13.4	601	4 US-09-949-016-176661	Sequence 176661, A
16	111.4	11.0	450	3 US-09-370-838-145	Sequence 145, App
17	111.4	11.0	450	3 US-09-854-133-145	Sequence 145, App
18	74	7.3	114793	4 US-10-148-806-3	Sequence 3, Appl
19	70.2	6.9	215	4 US-09-513-999C-31510	Sequence 31510, A
20	67	6.6	152132	4 US-09-949-016-13845	Sequence 13845, A
21	67	6.6	152145	4 US-09-949-016-12371	Sequence 12371, A
22	66.4	6.6	311	4 US-09-313-294A-5928	Sequence 5928, App
23	64.6	6.4	601	4 US-09-949-016-176640	Sequence 176640, A
24	64.6	6.4	601	4 US-09-949-016-176663	Sequence 176663, A
25	59.8	5.9	1505	1 US-07-915-246-1	Sequence 1, Appl
26	59	5.8	1225	4 US-09-976-594-416	Sequence 416, App
27	58.8	5.8	34230	4 US-09-949-016-12052	Sequence 12052, A

28	58.8	5.8	128470	4 US-09-949-016-13765	Sequence 13765, A
29	58.6	5.8	7218	1 US-08-232-463-14	Sequence 14, Appl
30	57.8	5.7	12695	4 US-09-949-016-16775	Sequence 16775, A
31	57	5.6	1926	3 US-09-249-585A-4	Sequence 4, Appl
32	57	5.6	1931	2 US-09-130-114-2	Sequence 2, Appl
33	55.4	5.5	1505	4 US-09-620-112D-544	Sequence 544, App
34	55.4	5.5	5682	4 US-10-164-595-1	Sequence 1, Appl
35	55.2	5.5	242	4 US-09-354-147C-30	Sequence 30, Appl
36	55.2	5.5	767677	4 US-09-949-016-12147	Sequence 12147, A
37	55.2	5.5	767677	4 US-09-949-016-17361	Sequence 17361, A
38	54.6	5.4	525	4 US-09-640-211A-177	Sequence 177, App
39	54.2	5.4	246	4 US-09-547-693-232	Sequence 232, App
40	54	5.3	1638	4 US-09-799-451-412	Sequence 412, App
41	53.6	5.3	324	4 US-09-547-693-234	Sequence 234, App
42	52.8	5.2	339	4 US-09-270-767-26436	Sequence 26436, A
43	52.8	5.2	3835	4 US-09-270-767-10943	Sequence 10943, A
44	51.4	5.1	472	1 US-08-361-467B-2	Sequence 2, Appl
45	51.4	5.1	472	1 US-08-484-332C-2	Sequence 2, Appl

## ALIGNMENTS

RESULT 1  
US-08-343-443B-1  
Sequence 1, Application US/0834343B  
Patent No. 5968734  
GENERAL INFORMATION:  
APPLICANT: Aurias, Alain  
APPLICANT: Delattre, Olivier  
APPLICANT: Desmarte, Chantal  
APPLICANT: Meiot, Thomas  
APPLICANT: Peter, Martine  
APPLICANT: Plougasatel, Beatrice  
APPLICANT: Thomas, Gilles  
APPLICANT: Zucman, Jessica  
TITLE OF INVENTION: NUCLEIC ACID CORRESPONDING TO A GENE OF  
TITLE OF INVENTION: CHROMOSOME 22 INVOLVED IN RECURRENT CHROMOSOMAL  
TITLE OF INVENTION: TRANSLATIONS ASSOCIATED WITH THE DEVELOPMENT OF CANCEROUS  
TITLE OF INVENTION: TUMORS, AND NUCLEIC ACIDS OF FUSION RESULTING FROM SAID  
TITLE OF INVENTION: TRANSLOCATIONS  
NUMBER OF SEQUENCES: 129  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Welser & Associates  
STREET: 230 South Fifteenth Street  
CITY: Philadelphia  
STATE: PA  
COUNTRY: USA  
ZIP: 19102  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: AEDIT 1.0 DOS text editor  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/343,443B  
FILING DATE: 18-NOV-1994  
CLASSIFICATION: 514  
PRIORITY APPLICATION DATA:  
PRIORITY APPLICATION NUMBER: PCT/FR93/00494  
FILING DATE: 19-MAY-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: FR 92/06123  
FILING DATE: 20-MAY-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Welser, Gerard J.  
REGISTRATION NUMBER: 19,763  
REFERENCE/DOCKET NUMBER: 989,6121P  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 215-875-8383  
TELEFAX: 215-875-8394  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:

```

/      LENGTH: 2371 base pairs
/
/      TYPE: nucleic acid
/
/      STRANDEDNESS: double
/
/      TOPOLOGY: linear
/
/      FEATURE:
/
/      NAME/KEY: CDS
/
/      LOCATION: 25..1992
/
US-08-343-443B-1

```

Query Match	99.8%	Score 1010.4;	DB 2;	Length 2371;
Best Local Similarity	99.9%	Pred. No. 2.8e-254;		
Matches 1011; Conservative	0;	Mismatches 1;	Indels 0;	Gaps 0;

OY	1	AGGAGCGCGGTGAATGCGGCACGCTCGAGAGACGAGGTGGCTTCAATAAGCCTGGTGGAC	60
Db	981	AGGAGCGCGGTGAATGCGGCACGCTCGAGAGACGAGGTGGCTTCAATAAGCCTGGTGGAC	1040
OY	61	CATGATGAAGGACCAAGTCTTTGATCTAGGCCACCTGTAGATCAGATGAAGCTCTGA	120
Db	1041	CATGATGAAGGACCAAGTCTTTGATCTAGGCCACCTGTAGATCAGATGAAGCTCTGA	1100
OY	121	CAACGATGCAATTTATGTACAAAGATTAATGACAGGTGACTCTAGATGATCTGGCAGA	180
Db	1101	CAACGATGCAATTTATGTACAAAGATTTAAATGACAGGTGACTCTAGATGATCTGGCAGA	1160
OY	181	CTTCTTTAAGCAGTGTGGGGTTGTTAAGATGAACAAGAACTGGGCAATCCATGATCCA	240
Db	1161	CTTCTTTAAGCAGTGTGGGGTTGTTAAGATGAACAAGAACTGGGCAATCCATGATCCA	1220
OY	241	CATCTAAGCTGGAACAAGAAACAAGAAACCCAAAGGCGATGCCACAGTCTATGAA	300
Db	1221	CATCTAAGCTGGAACAAGAAACAAGAAACCCAAAGGCGATGCCACAGTCTATGAA	1280
OY	301	CCCAACCACTGCGCAAGGCTGCGGTGGAATGTTGATGGGAAGATTTTCAAGGAGCAA	360
Db	1281	CCCAACCACTGCGCAAGGCTGCGGTGGAATGTTGATGGGAAGATTTTCAAGGAGCAA	1340
OY	361	ACTTAAAGTCTCCCTTGGCTCGGAAGAACCTTCCATGAACGATGCGGGGTGGTCTGGC	420
Db	1341	ACTTAAAGTCTCCCTTGGCTCGGAAGAACCTTCCATGAACGATGCGGGGTGGTCTGGC	1400
OY	421	ACCCGCTGAGGGCAGAAGGCATGCCACAACACTCGGTGAAGGTCAAGAGGCCCAAGAGG	480
Db	1401	ACCCGCTGAGGGCAGAAGGCATGCCACAACACTCGGTGAAGGTCAAGAGGCCCAAGAGG	1460
OY	481	TCCTGGGGGAGCCCATGGGTGCGATGGGAGGCCGTGGAGAGATGAGGAGCTTCCCTCC	540
Db	1461	TCCTGGGGGAGCCCATGGGTGCGATGGGAGGCCGTGGAGAGATGAGGAGCTTCCCTCC	1520
OY	541	AAGAAGAACCCCGGGGTTCCCGAGGGGAAACCCCTCTGGAAGAGAGAAACGTCCAGACCGAGC	600
Db	1521	AAGAAGAACCCCGGGGTTCCCGAGGGGAAACCCCTCTGGAAGAGAGAAACGTCCAGACCGAGC	1580
OY	601	TGGAGACTGGCAGTGTCCCAATCCGGGTGTGGAAACCAAGACTTCGCTGGAGAACGA	660
Db	1581	TGGAGACTGGCAGTGTCCCAATCCGGGTGTGGAAACCAAGACTTCGCTGGAGAACGA	1640
OY	661	GTGCAACAAGTGTAAAGGCCCAAAACCTGGAAGGCTTCTCCGCCAACCTTCCGCCCCC	720
Db	1641	GTGCAACAAGTGTAAAGGCCCAAAACCTGGAAGGCTTCTCCGCCAACCTTCCGCCCCC	1700
OY	721	GGGTGATGATCTGTGCACAGATGTGCGCTGTGTGCATGCGGGGAGGAAGGTGGCTCAT	780
Db	1701	GGGTGATGATCTGTGCACAGATGTGCGCTGTGTGCATGCGGGGAGGAAGGTGGCTCAT	1760
OY	781	GGATGTGTGGTCCCGGTGGAATGTTCAAGGTGGCGGTGTGAGACAGAGTGGCTT	840
Db	1761	GGATGTGTGGTCCCGGTGGAATGTTCAAGGTGGCGGTGTGAGACAGAGTGGCTT	1820
OY	841	CGGTGGTGGCGGGGCAATGGAACGAGGTGGCTTGTGTGAAGGAACAAGGTGGCGCTGG	900
Db	1821	CGGTGGTGGCGGGGCAATGGAACGAGGTGGCTTGTGTGAAGGAACAAGGTGGCGCTGG	1880

Qy	901	GGGGCCCCCTGGACCTTTGATGGAA	CAGATGGGGGAAAGAAAGSAGAGACG	GTGGAGAGCC	960
Db	1881	GGGGCCCCCTGGACCTTTGATGGAA <td>CAGATGGGGGAAAGAAAGSAGAGACG</td> <td>COTGGAGAGCC</td> <td>1940</td>	CAGATGGGGGAAAGAAAGSAGAGACG	COTGGAGAGCC	1940
Qy	961	TGGAAAAATGGATTAAGCGAGACCGCT	CAAGCGGAGAGATGGGCTTAC		1012
Db	1941	TGGAAAAATGGATTAAGCGAGACCGCT	CAAGCGGAGAGATGGGCTTAC		1992

RESULT 2  
US-09-94

; Sequence 5043, Application US/09949016  
Patent No. 6813238

; PATENT NO. 0812555  
; GENERAL INFORMATION  
; GENERAL INFORMATION

APPLICANT: VENTER, J. Craig et al.  
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED

FILE REFERENCE: CL001307

CURRENT APPLICATION NUMBER: US/09/949,016  
CURRENT FILING DATE: 2000-04-14

PRIOR APPLICATION NUMBER: 60/241,755

;  
PRIOR FILING DATE: 2000-10-20  
;  
PRIOR APPLICATION NUMBER: 60/237,768  
;

PRIOR FILING DATE: 2000-10-03  
PRIOR APPLICATION NUMBER: 60/231,498

PRIOR FILING DATE: 2000-09-08  
NUMBER OF SEQ ID NOS: 207012

```

; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0.0

```

```

; SEQ ID NO 5043
; LENGTH: 1785

```

TYPE: DNA  
ORGANISM: Human

US-09-949-016-5043

Query Match	Score	DB 4	Length
94.5%	956	1785	

Best Local Similarity 100.0%; Pred. NO. 4,26-240;  
Matches 956; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

57 GACCCATGATGAAGGACCAAGTCTTGATCTAGGCCACCTGTAGATCCAGATGAAGACT 116

[illegible]

התאריך: 20/01/2020

117 CTGACACAGTGCATTATGACAGGATTAAATGACAGTGTGATCTAGATGATCTGG 1/6

Db 887 CTGACACAGTGCATTTATGTACACAGATTAAATGCACAGTGTGACTCTAGATGATCTGG 946

177 CAGACTTCTTAAGCAGTGGGGTTGTTAAGATGAACAAGAACTGGGCAACCATGA 236

947 CAGACTTCTTTAAGCAGTGTGGGTTGTTAAGATGACAAGAGAACTGGGCAACCCATGA 1006

237 TCACATCTACCTGGACAAGAAACAGAAAGCCCAAGCGATGCCACAGTGTCTTATG 296

Db 1007 TCCACATCTACCTGGACAGGAAACAGGAAAGCCGATGCCACAGTGTCCATTAG 1066

307 **ANALYSIS OF THE EFFECTS OF THE 1997-1998 EL NIÑO ON THE RAINFALL AND TEMPERATURE IN THE TROPICAL AMERICAS** 356

[illegible]

Db 1067 AAGACCCACCTGCCAAGGCTGCCGTGGAAAGTGTGATGGGAAGATTTTCAAGGA 1126

357 GCAACTTAAAGTCTCCCTTGCTCGAGAGAGCTCCATGAAACAGTATCGGGGTGTC 416

Db 1127 GCAACTTAAAGTCTCCCTTGTGCGAAGAGCCTCCATGAACAGTATGCGGGGTGTC 1186

417 TGCACCCCGTAGGGGAGAGGCATGCCACCACTCCGTGAGGTTCCAGAGGCCAG 476

Db 1187 TGCCACCCCGTAGGGCAGAGCATGCCACCACTCCGTGAGGTCCAGAGGCCAG 1246

477 GAGTCTCTGGGGAACCCATGGTGCATGGGAAGCGGTGAGGAGATTAGAGAGGCTTCC 536

[illegible][illegible]

537 CTCCAGAGAACCCCGGGTTCCTCCGAGGGAACCCCTCTGAGAGAGAAACGTCCAGACC 596

Db 1307 CTCGAAGAGGACCCCGGGGTTCCGAGGGAAACCCCTTGAGAGAGGAAACGTCCAGCAC 1366

```
QY 597 GAGCTGAGACAGTGGCAATGTCCTCAATCCGGGTTGTGAAACAGAACTTGCTGAGAA 656
D 1367 GAGCTGAGACAGTGGCAATGTCCTCAATCCGGGTTGTGAAACAGAACTTGCTGAGAA 1426
QY 657 CAGAGTCAACCAAGTGAAGGCCCCAAGGCTGAAAGGCTTCCTCCGCAACCTTTCGCG 716
D 1427 CAGAGTCAACCAAGTGAAGGCCCCAAGGCTGAAAGGCTTCCTCCGCAACCTTTCGCG 1486
QY 717 CCCCAGGTGTGATCTGTGAGAGAGTGGCCCTGTGTGCATGCGGGGAGAGAGAGTGGCC 776
D 1487 CCCCAGGTGTGATCTGTGAGAGAGTGGCCCTGTGTGCATGCGGGGAGAGAGAGTGGCC 1546
QY 777 TCATGATCTGTGTGTGTCCTGGTGAATGTTCAAGAGTGGCCCTGTGTGAGAGAGAGTGTG 836
D 1547 TCATGATCTGTGTGTGTCCTGGTGAATGTTCAAGAGTGGCCCTGTGTGAGAGAGAGTGTG 1606
QY 837 GCTTCGCTGTGTGCGCGGGGAGATGAGACGAGGTGGCTTGTGTGAGAGAGAGAGTGGCC 896
D 1607 GCTTCGCTGTGTGCGCGGGGAGATGAGACGAGGTGGCTTGTGTGAGAGAGAGAGTGGCC 1666
QY 897 CTGGGGGGCCCCCTGAGACTTTGATGAGAACAGATGGAGAGAGAGAGAGAGAGTGGAG 956
D 1667 CTGGGGGGCCCCCTGAGACTTTGATGAGAACAGATGGAGAGAGAGAGAGAGAGTGGAG 1726
QY 957 GACCTGAGAAAATGATTAAGGAGAGACCGCTCAGAGCGCAGAGATCGGCCCTTAC 1012
D 1727 GACCTGAGAAAATGATTAAGGAGAGACCGCTCAGAGCGCAGAGATCGGCCCTTAC 1782
```

## RESULT 3

```
US-09-949-016-5044
; Sequence 5044, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTNER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: C1001307
; CURRENT APPLICATION NUMBER: US/09/949, 016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 5044
; LENGTH: 1785
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-5044
```

```
Query Match 94.5%; Score 956; DB 4; Length 1785;
Best Local Similarity 100.0%; Pred. No. 4.2e-240;
Matches 956; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 57 GACCCATGATGTAAGGACGAGATCTTATCTAGCCCACTGTAGATCCAGATGAAGACT 116
D 827 GACCCATGATGTAAGGACGAGATCTTATCTAGCCCACTGTAGATCCAGATGAAGACT 886
QY 117 CTGACAACAGTGCATTTATGATCAAGATTAATGACAGTGTGACTCTAGATGATCTG 176
D 887 CTGACAACAGTGCATTTATGATCAAGATTAATGACAGTGTGACTCTAGATGATCTG 946
QY 177 CAGACTCTTTAAGCAGTGTGGGTTGTTAAGATGAACAAGAACTGGCCAAACCATGA 236
D 947 CAGACTCTTTAAGCAGTGTGGGTTGTTAAGATGAACAAGAACTGGCCAAACCATGA 1006
QY 237 TCACATCTACCTGGAACAAGAAACAGAAAGCCCAAGGCGATGCCAGTGTCCATG 296
```

```
D 1007 TCACATCTACCTGGAACAAGAAACAGAAAGCCCAAGGCGATGCCAGTGTCTATG 1066
QY 297 AAGACCAACCACTGTGCAAGGCTGCGGTGGAATGTTTATGGAAGATTTTCAAGGA 356
D 1067 AAGACCAACCACTGTGCAAGGCTGCGGTGGAATGTTTATGGAAGATTTTCAAGGA 1126
QY 357 GCAACTTAATAGTCTCCCTTGTCTGCAAGAGAGTCCCAAGAACATGATCCGGGTGTC 416
D 1127 GCAACTTAATAGTCTCCCTTGTCTGCAAGAGAGTCCCAAGAACATGATCCGGGTGTC 1186
QY 417 TGCACACCCGTGAGGGGAGAGGATGCAACCACTCCGTGAGAGTCCAGAGAGCCAG 476
D 1187 TGCACACCCGTGAGGGGAGAGGATGCAACCACTCCGTGAGAGTCCAGAGAGCCAG 1246
QY 477 GAGTCTGTGGGAGACCCATGAGTGCATGAGAGGCGGTGAGAGAGATGAGAGGCTTCC 536
D 1247 GAGTCTGTGGGAGACCCATGAGTGCATGAGAGGCGGTGAGAGAGATGAGAGGCTTCC 1306
QY 537 CTCCAAGAGACCCCGGGGTTCCCGAGAGAACCCCTCTGAGAGAGAGAAAGTCCAGACC 596
D 1307 CTCCAAGAGACCCCGGGGTTCCCGAGAGAACCCCTCTGAGAGAGAGAAAGTCCAGACC 1366
QY 597 GAGCTGAGACAGTGAAGTCCCAATCCGGGTTGTGAAACAGAACTTGCGCTGAGAA 656
D 1367 GAGCTGAGACAGTGAAGTCCCAATCCGGGTTGTGAAACAGAACTTGCGCTGAGAA 1426
QY 657 CAGAGTCAACCAAGTGAAGGCCCCAAGGCTTGAAGGCTTCTCCGCAACCTTTCGCG 716
D 1427 CAGAGTCAACCAAGTGAAGGCCCCAAGGCTTGAAGGCTTCTCCGCAACCTTTCGCG 1486
QY 717 CCCCAGGTGTGATCTGTGAGAGAGTGGCCCTGTGTGCATGCGGGGAGAGAGAGTGGCC 776
D 1487 CCCCAGGTGTGATCTGTGAGAGAGTGGCCCTGTGTGCATGCGGGGAGAGAGAGTGGCC 1546
QY 777 TCATGATCTGTGTGTGTCCTGGTGAATGTTCAAGAGTGGCCCTGTGTGAGAGAGAGTGTG 836
D 1547 TCATGATCTGTGTGTGTCCTGGTGAATGTTCAAGAGTGGCCCTGTGTGAGAGAGAGTGTG 1606
QY 837 GCTTCGCTGTGTGCGCGGGGAGATGAGACGAGGTGGCTTGTGTGAGAGAGAGAGTGGCC 896
D 1607 GCTTCGCTGTGTGCGCGGGGAGATGAGACGAGGTGGCTTGTGTGAGAGAGAGAGTGGCC 1666
QY 897 CTGGGGGGCCCCCTGAGACTTTGATGAGAACAGATGGAGAGAGAGAGAGAGTGGAG 956
D 1667 CTGGGGGGCCCCCTGAGACTTTGATGAGAACAGATGGAGAGAGAGAGAGAGTGGAG 1726
QY 957 GACCTGAGAAAATGATTAAGGAGAGACCGCTCAGAGCGCAGAGATCGGCCCTTAC 1012
D 1727 GACCTGAGAAAATGATTAAGGAGAGACCGCTCAGAGCGCAGAGATCGGCCCTTAC 1782
```

## RESULT 4

```
US-09-949-016-13696
; Sequence 13696, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTNER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: C1001307
; CURRENT APPLICATION NUMBER: US/09/949, 016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 13696
; LENGTH: 6002
; TYPE: DNA
```

ORGANISM: Human  
US-09-949-016-13696

Query Match 90.0%; Score 911.2; DB 4; Length 6002;  
Best Local Similarity 95.1%; Pred. No. 3.8e-228;  
Matches 962; Conservative 0; Mismatches 48; Indels 2; Gaps 2;

1 AGAAGCGGTGGTAAATGGGCAAGCCCTGGAGACCAAGGTGGCTTCAATAGCTGTGGAGACC 60  
2990 AGAATGGCGGTGGAATGGGCAAGCCCTGGAGACCAAGGTGGCTTCAATAGCTGTGGAGACC 3049  
61 CATGATGAAGAAGCAAGATCTTGATCTAGAGCCCACTGTAGATCCAGATGAAGACTCTGA 120  
3050 CATGATGAAGAAGCAAGATCTTGATCTAGAGCCCACTGTAGATCCAGATGAAGACTCTGA 3109  
121 CAACAGTGAATTTATGTACAAAGATTAAATGACAGTGTGACTGTAGATATCTGGAGCA 180  
3110 CAACAGTGAATTTATGTACAAAGATTAAATGACAGTGTGACTGTAGATATCTGGAGCA 3169  
181 CTTCTTTAAGCAATGTTGGGTTGTTAGATGAACAAAGAACTGGGCAACCAATGATCCA 240  
3170 CTTCTTTAAGCAATGTTGGGTTGTTAGATGAACAAAGAACTGGGCAACCAATGATCCA 3229  
241 CATCTACCTGGAACAAGAAACAGGAAGCCCAAGGCGATGCCAAGTGTCTATGAGA 300  
3230 CACTTACTGGAACAAGAAACAGGAAGCCCAAGGCGATGCCAAGTGTCTATGAGA 3289  
301 CCCAACCCTGGAACAAGGCTGCGTGGATGTTGATGGAGAAAGATTTTCAAGGAGCA 360  
3290 CTCACTTACTGGAACAAGGCTGCGTGGATGTTGATGGAGAAAGATTTTCAAGGAGCA 3349  
361 ACTTAAAGTCTCCCTGCTGGAAGAGCCCTCAAGTGAACGATGAGGGGTGGTCTGCC 420  
3350 ACTTAAAGTCTCCCTGCTGGAAGAGCCCTCAAGTGAACGATGAGGGGTGGTCTGCC 3409  
421 ACCCGTGAAGGAGCAAGGATGCAACCACTCCGTGAGTCCAGAGGCGCCAGAGG 480  
3410 ACCCGTGAAGGAGCAAGGATGCAACCACTCCGTGAGTCCAGAGGCGCCAGAGG 3469  
481 TCCTGGGGGAGCCCATGGGTGCAATGGAGGCGGTGAGAGAGATGAGAGGCTTCCCTCC 540  
3470 TCCTGGGGGAGCCCATGGGTGCAATGGAGGCGGTGAGAGAGATGAGAGGCTTCCCTCC 3529  
541 AAGAGGACCCCGGGGTTCCCGAGGGAACCCCTGTGAGAGAGAAACCTTCCAGCAACGAGC 600  
3530 AAGAGGACCCCGGGGTTCCCGAGGGAACCCCTGTGAGAGAGAAACCTTCCAGCAACGAGC 3589  
601 TGGAGACTGGCAGTGTCCCAATCCGGGTGTGAGAAACAGAACTTCCCTGAGAGACAGA 660  
3590 TGGAGACTGGCAGTGTCCCAATCCGGGTGTGAGAAACAGAACTTCCCTGAGAGACAGA 3649  
661 GTGCAACAGTGAAGGCCCAAGCCCTGAAGGCTTCCCTCCGACACCTTTCCGCCCC 720  
3650 GAGCAACAAGTGAAGGCCCTCAAGGCTTGAAGGCTTCCCTCCGACACCTTTCCGCCCC 3709  
721 GGGTGTGATCTGATGAGAGGTTGGGCTGTGAGGAAATCCGGGAGAGAAAGAGTGGCTCAT 780  
3710 GGGTGTGATCTGATGAGAGGTTGGGCTGTGAGGAAATCCGGGAGAGAAAGAGTGGCTCAT 3769  
781 GGATCGTGTGATCTCCGGTGAAGTTTCAAGAGTGGCGTGTGAGAGACAGAGGTGGCTT 840  
3770 GGATCGTGTGATCTCCGGTGAAGTTTCAAGAGTGGCGTGTGAGAGACAGAGGTGGCTT 3829  
841 CCGTGTGAGCCGGGAGCATGACCCAGAGGTGCTTTGTGTGAGAGAAAGCAGAGTGGCTTGG 900  
3830 CCGTGTGAGCCGGGAGCATGACCCAGAGGTGCTTTGTGTGAGAGAAAGCAGAGTGGCTTGG 3888  
901 GGGGCCCCCTGGAACCTTTGATGAACAGATGGGAGAGAAAGAGAGAGAGCTGAGAGACC 960  
3889 GGGGCCCCCTGGAACCTTTGATGAACAGATGGGAGAGAAAGAGAGAGAGCTGAGAGACC 3947  
961 TGGAAAAATGATTAAGGCGAGCAACCGTCAAGAGCGAGAGATGGGCTTAC 1012

Db 3948 TGGAAAAATGATTAAGGCGAGCACTGTCAAGAGCGCAGAGATCAGCCCTAC 3999

# RESULT 5

US-09-949-016-1954  
Sequence 1954, Application US/09949016  
Patent No. 6812339  
GENERAL INFORMATION:  
APPLICANT: VENTER, J. Craig et al.  
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
FILE REFERENCE: CL001307  
CURRENT APPLICATION NUMBER: US/09/949,016  
CURRENT FILING DATE: 2000-04-14  
PRIOR APPLICATION NUMBER: 60/241,755  
PRIOR FILING DATE: 2000-10-20  
PRIOR APPLICATION NUMBER: 60/237,768  
PRIOR FILING DATE: 2000-10-03  
PRIOR APPLICATION NUMBER: 60/231,498  
PRIOR FILING DATE: 2000-09-08  
NUMBER OF SEQ ID NOS: 207012  
SOFTWARE: FASTSEQ for Windows Version 4.0  
SEQ ID NO 1954  
LENGTH: 1783  
TYPE: DNA  
ORGANISM: Human  
US-09-949-016-1954

Query Match 85.0%; Score 860; DB 4; Length 1783;  
Best Local Similarity 95.1%; Pred. No. 5.5e-215;  
Matches 909; Conservative 0; Mismatches 45; Indels 2; Gaps 2;

57 GACCCATGATGAAGAGCAAGATCTGATCTAGGCCCACTGTAGATCCAGATGAAGACT 116  
827 GACCCATGATGAAGAGCAAGATCTGATCTAGGCCCACTGTAGATCCAGATGAAGACT 886  
117 CTGACAACAGTGAATTTATGTACAAAGATTAAATGACAGTGTGACTGTAGATGATCTGG 176  
887 CTGACAACAGTGAATTTATGTACAAAGATTAAATGACAGTGTGACTGTAGATGATCTGG 946  
177 CAGACTTTCTTAAAGCAGTGTGGGTTGTTAAGTGAACAAAGAACTGGGCAACCAATGA 236  
947 TAGACTTTCTTAAAGCAGTGTGGGTTGTTAAGTGAACAAAGAACTGGGCAACCAATGA 1006  
237 TCCACATTAACCTGGAACAAGAAACAGAAAGCCCAAGGCGATGCAAGTGTCTATG 236  
1007 TCCACATTAACCTGGAACAAGAAACAGAAAGCCCAAGGCGATGCAAGTGTCTATG 1066  
297 AAGACCAACCACTGCAAGAGGCTGCGTGAATGTTTGAAGGAAAGATTTTCAAGGGA 356  
1067 AAGACCAACCACTGCAAGAGGCTGCGTGAATGTTTGAAGGAAAGATTTTCAAGGGA 1126  
357 GCAAACTTAAAGTCTCCCTTGTCTGGAAGAAAGCTTCAATGAACAGTATCGGGGTGTC 416  
1127 GCAAACTTAAAGTCTCCCTTGTCTGGAAGAAAGCTTCAATGAACAGTATCGGGGTGTC 1186  
417 TGGCAACCCCGTGAAGGCAAGGATGCAACCACTCCGTGAGAGTCCAGAGGCGCCAG 476  
1187 TGGCAACCCCGTGAAGGCAAGGATGCAACCACTCCGTGAGAGTCCAGAGGCGCCAG 1246  
477 GAGGTCTGTGGGAGCAACATGGGTGCAATGGAGAGGCGGTGAGAGAGATGAGAGGCTTCC 536  
1247 GAGGTCTGTGGGAGCAACATGGGTGCAATGGAGAGGCGGTGAGAGAGATGAGAGGCTTCC 1306  
537 CTCGAAGAGACCCCGGGGTTCCGAGGGAACCTCTGTGAGAGAGAAACGTTCCAGACAC 596  
1307 CTCGAAGAGACCCCGGGGTTCCGAGGGAACCTCTGTGAGAGAGAAACGTTCCAGACAC 1366  
597 GAGCTGGAACCTGCAAGTGTCCCAATCCGGGTGTGAGAAACAGAACTTGGCTGAGAGAA 656  
1367 GAGCTGGAACCTGCAAGTGTCCCAATCCGGGTGTGAGAAACAGAACTTGGCTGAGAGAA 1426  
657 CAGAGTGAACAGTGTAAAGGCCCAAGGCTGAAGGCTTCTCCGCGCACCTTTCCGC 716

Db	1427	CAGAGGCAACAAGTGTAAAGCTTCAAAGCCTTGAAGCTTCTCCCGCACCTTTCCAC	1486
Oy	717	CCCCGGATGTGATATGTGGCAGAAGGTGGCCCTTGATGCATGCGGGAGAAAGAGGTGCC	776
Db	1487	CCCCGGATGTGATATGTGGCAGAAGGTGGCCCTTGATGCATGCGGGAGAAAGAGGTGCC	1546
Oy	777	TCAATGAATCGTGATGTGTCCCGGTGAAATGTTCAAGATGTGGCCCTGTGTGAGACAAGAGTG	836
Db	1547	TCAATGAATCATGTGTGTCTCCGGTGAATGTTCAAGATGTGGCCCTGTGTGTGAGACAAGAGTG	1606
Oy	837	GCTTCCGTGTGTGCGCGGSCAATGAGCCGAGGTGCTTTTGATGTAAGAAACAAGGTGGCC	896
Db	1607	GCTTCCGTGTGTGCG - CTGGGCAATGAGCCGAGGTGCTTTTGATGTAAGAAACAAGGTGGCC	1666
Oy	897	CTGGGGGGCCCCCTTGACCTTTGATGAAACAGATGGAGAGAAAGAGAGAGAGACTGGAG	956
Db	1666	CTGGGGGGGGCCCC - GGACCTTTGATGTAACCAATGAGGAGAAAGAAAGAGAGAGACTGGANG	1724
Oy	957	GACCTGAAAAATGATAAAGGCGAGCAACCGTCAGAGCGCAGAGATAGGGCCCTAC	1012
Db	1725	GACCTGAAAAACGATATTAAAGCGACACTGTTCAGAGGCGCAGAGATACGCCCTAC	1780

## RESULT 6

```

US-09-621-976-13361
Sequence 13361, Application US/09621976
Patent No. 6639063
GENERAL INFORMATION:
APPLICANT: Dumas Milne Edwards, J.B.
APPLICANT: Jobert, S.
APPLICANT: Giordano, J.Y.
TITLE OF INVENTION: ESTs and Encoded Human Proteins
FILE REFERENCE: GENSET .054PR2
CURRENT APPLICATION NUMBER: US/09/621,976
CURRENT FILING DATE: 2000-07-21
NUMBER OF SEQ ID NOS: 19335
SOFTWARE: Patent.pm
SEQ ID NO 13361
LENGTH: 411
TYPE: DNA
ORGANISM: Homo sapiens
US-09-621-976-13361

```

Query Match 33.4%; Score 338; DB 4; Length 411;

Best Local Similarity 99.7%; Pred. No. 1.1e-78;  
Matches 349; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

Qy	56	UGACCCATGATGTAAGAACCAAGATTTGATCTAGGCCCACTGTAGATCCAGATGAGAC	115
Db	62	GGAACCAATGATGTAAGGACCAAGATCTTATCTAGGCCCACTGTAGATCCAGATGAGAC	121
Qy	116	TCTGCACAAGGCAATTTATGTACAAGGATTTAAATGACAGTGACCTGTAGATGATCTG	175
Db	122	TCTGCACAAGGCAATTTATGTACAAGGATTTAAATGACAGTGACCTGTAGATGATCTG	181
Qy	176	GCAGCTCTTTAAGCAGCTGTGGGTTGTTAAGATGAACAAGACCTGGCAACCATG	235
Db	182	GCAGCTCTTTAAGCAGCTGTGGGTTGTTAAGATGAACAAGACCTGGCAACCATG	241
Qy	236	ATCCACATCTACTGGAACAAGGAAACAGAAAGCCCAAGGCGA-TGCCACAGTGTCTA	294
Db	242	ATCCACATCTACTGGAACAAGGAAACAGAAAGCCCAAGGCGATGCCACAGTGTCTA	301
Qy	295	TGAAGACCCACCCACTGCCAAGGCTGCCGTGGAATGTTGATGGAAGAATTTCAAGG	354
Db	302	TGAAGACCCACCCACTGCCAAGGCTGCCGTGGAATGTTGATGGAAGAATTTCAAGG	361
Qy	355	GAGCAAACTTAAAGTCTCCCTTGTCCGGAAGAAGCTCCAAATGAACAATA	404
Db	362	GAGCAAACTTAAAGTCTCCCTTGTCCGGAAGAAGCTCCAAATGAACAATA	411

```

RESULT 7
US-09-949-016-16785
; Sequence 16785, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949, 016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: fastseq for Windows Version 4.0
; SEQ ID NO 16785
; LENGTH: 35784
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-16785

```

## Best I

	Matches 268;	Conservative 0;	Mismatches 16;	Indels 0;	Gaps 0;
QY	692	GGCTTCTCCCGCACCCCTTTCGCCGCCGGGNGGATCGGAGAGAGTGCCCTGGT	751		
Db	33261	GACTGCTTTTGCCCTCTGTAATTCATCCTTAGTGATGATCGACAGAGTGCCCTGGT	33320		
QY	752	GGCATGCGGGAGAGAGAGGTGGCTCATGATCGTGATGTCCTCGGTGAAATGTTACAG	811		
Db	33321	GGCATGCGGGAGAGAGAGGTGGCTCATGATCGTGATGTCCTCGGTGAAATGTTACAG	33380		
QY	812	GGTGGCCGTGTGTGAGACAGAGGTGGCTTCGTGTGGCGCGGGATGAGCCGAGGTGGC	871		
Db	33381	GGTGGCCGTGTGTGAGACAGAGGTGGCTTCGTGTGGCGGGAGATGAGCCGAGGTGGC	33440		
QY	872	TTTGTGTGAGAGAGCAGAGGTGGCCCTGTGGGGGCCCCCTTGACCTTTTATGTGAAACAGATG	931		
Db	33441	TTTGTGTGAGAGAGCAGAGGTGGCCCTGTGGGGGCCCCCTTGACCTTTTATGTGAAACAGATG	33500		
QY	932	GGAGAGAAAGAGAGAGACTGTGAGGACTCTGAAAAATGTGATTA	975		
Db	33501	GGAGAGAAAGAGAGAGACTGTGAGGACTCTGAAAAATGTGATTA	33544		

## RESULT &

```

US-09-949-016-16786
Sequence 16786, Application US/09949016
Patent No. 6812339
GENERAL INFORMATION:
APPLICANT: VENTER, J. Craig et al.
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
FILE REFERENCE: CLO01307
CURRENT APPLICATION NUMBER: US/09/949,016
CURRENT FILING DATE: 2000-04-14
PRIOR APPLICATION NUMBER: 60/241,755
PRIOR FILING DATE: 2000-10-20
PRIOR APPLICATION NUMBER: 60/237,768
PRIOR FILING DATE: 2000-10-03
PRIOR APPLICATION NUMBER: 60/231,498
PRIOR FILING DATE: 2000-09-08
NUMBER OF SEQ ID NOS: 207012
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 16786
LENGTH: 35784
TYPE: DNA
ORGANISM: Human

```







US-09-949-016-176660  
; Sequence 176660, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: C1001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FASTSEQ for Windows Version 4.0  
; SEQ ID NO 176660  
; LENGTH: 601  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-176660

Query Match 13.4%; Score 135.8; DB 4; Length 601;  
Best Local Similarity 98.6%; Pred. No. 1e-25;  
Matches 137; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 201 TTGTTAAGATGAACAGAGAACTGGGCAACCATGATCCATCTTACCTGGACAGGAAA 260  
DB 34 TTGCTAGATGAACAGAGAACTGGGCAACCATGATCCATCTTACCTGGACAGGAAA 93  
QY 261 CAGAAAGCCCAAGGCGATGCCACATGTCCTATGAAGACCCCACTGCCAAGGCTG 320  
DB 94 CAGAAAGCCCAAGGCGATGCCACATGTCCTATGAAGACCCCACTGCCAAGGCTG 153  
QY 321 CCGTGAATGTTTGATGG 339  
DB 154 CCGTGAATGTTTGATGG 172

RESULT 15  
US-09-949-016-176661  
; Sequence 176661, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: C1001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FASTSEQ for Windows Version 4.0  
; SEQ ID NO 176661  
; LENGTH: 601  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-176661

Query Match 13.4%; Score 135.8; DB 4; Length 601;  
Best Local Similarity 98.6%; Pred. No. 1e-25;  
Matches 137; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 201 TTGTTAAGATGAACAGAGAACTGGGCAACCATGATCCATCTTACCTGGACAGGAAA 260  
DB 25 TTGCTAGATGAACAGAGAACTGGGCAACCATGATCCATCTTACCTGGACAGGAAA 84

QY 261 CAGAAAGCCCAAGGCGATGCCACATGTCCTATGAAGACCCCACTGCCAAGGCTG 320  
DB 85 CAGAAAGCCCAAGGCGATGCCACATGTCCTATGAAGACCCCACTGCCAAGGCTG 144  
QY 321 CCGTGAATGTTTGATGG 339  
DB 145 CCGTGAATGTTTGATGG 163

Search completed: February 20, 2005, 16:33:17  
Job time: 177.828 secs



LENGTH: 1988  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-10-094-749-1074

Query Match 100.0%; Score 1012; DB 17; Length 1988;  
Best Local Similarity 100.0%; Pred. No. 1.5e-283;  
Matches 1012; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 AGGACGGGTGGAATGGGCGCTGGAGCGAGGTGGCTTCAATAGCTGTGGACC 60
DB AGGACGGGTGGAATGGGCGCTGGAGCGAGGTGGCTTCAATAGCTGTGGACC 859
QY 61 CATGATGAAGACCAAGATCTTATCTAGGCCCCCTGTAGATCCAGATGAACCTCTGA 120
DB CATGATGAAGACCAAGATCTTATCTAGGCCCCCTGTAGATCCAGATGAACCTCTGA 919
QY 121 CAACAGTGAATTTATATACAGGATTTAAATGACAGTGTACTCTAGATGATCTGGCAGA 180
DB CAACAGTGAATTTATATACAGGATTTAAATGACAGTGTACTCTAGATGATCTGGCAGA 979
QY 920 CAACAGTGAATTTATATACAGGATTTAAATGACAGTGTACTCTAGATGATCTGGCAGA 979
DB CAACAGTGAATTTATATACAGGATTTAAATGACAGTGTACTCTAGATGATCTGGCAGA 1039
QY 181 CTCTTTAAGCAAGTGTGGGTTGTTAAGTGAACAGAGAACTGGGCAACCATGATCCA 240
DB CTCTTTAAGCAAGTGTGGGTTGTTAAGTGAACAGAGAACTGGGCAACCATGATCCA 1039
QY 980 CTCTTTAAGCAAGTGTGGGTTGTTAAGTGAACAGAGAACTGGGCAACCATGATCCA 1039
DB CTCTTTAAGCAAGTGTGGGTTGTTAAGTGAACAGAGAACTGGGCAACCATGATCCA 1099
QY 241 CATCTACCTGGAACAGAGAAACAGAGAAACCCAAAGGCGATGCCAGTGTCTATGAGA 300
DB CATCTACCTGGAACAGAGAAACAGAGAAACCCAAAGGCGATGCCAGTGTCTATGAGA 1099
QY 1040 CATCTACCTGGAACAGAGAAACAGAGAAACCCAAAGGCGATGCCAGTGTCTATGAGA 1099
DB CATCTACCTGGAACAGAGAAACAGAGAAACCCAAAGGCGATGCCAGTGTCTATGAGA 1159
QY 301 CCCACCACTGCAAGGCTGCGTGAATGTTGATGGGAAAGATTTCAAGGAGCA 360
DB CCCACCACTGCAAGGCTGCGTGAATGTTGATGGGAAAGATTTCAAGGAGCA 1159
QY 1100 CCCACCACTGCAAGGCTGCGTGAATGTTGATGGGAAAGATTTCAAGGAGCA 1159
DB CCCACCACTGCAAGGCTGCGTGAATGTTGATGGGAAAGATTTCAAGGAGCA 1219
QY 361 ACTTAAAGTCTCCCTTGTGCGAAGAACCTTCAATGAACATATGCGGGGTGTCTGCC 420
DB ACTTAAAGTCTCCCTTGTGCGAAGAACCTTCAATGAACATATGCGGGGTGTCTGCC 1219
QY 1160 ACTTAAAGTCTCCCTTGTGCGAAGAACCTTCAATGAACATATGCGGGGTGTCTGCC 1219
DB ACTTAAAGTCTCCCTTGTGCGAAGAACCTTCAATGAACATATGCGGGGTGTCTGCC 1279
QY 421 ACCCGTGAAGGAGAGGATGCAACCACTCCGTGAGAGTCAAGAGGCCCGAGAGG 480
DB ACCCGTGAAGGAGAGGATGCAACCACTCCGTGAGAGTCAAGAGGCCCGAGAGG 1279
QY 1220 ACCCGTGAAGGAGAGGATGCAACCACTCCGTGAGAGTCAAGAGGCCCGAGAGG 1279
DB ACCCGTGAAGGAGAGGATGCAACCACTCCGTGAGAGTCAAGAGGCCCGAGAGG 540
QY 481 TCCGTGGGGGAGCCCATGGGTGCGATGGAGGCGGTGAGAGAGATAGAGAGGCTTCCCTCC 540
DB TCCGTGGGGGAGCCCATGGGTGCGATGGAGGCGGTGAGAGAGATAGAGAGGCTTCCCTCC 1339
QY 1280 TCCGTGGGGGAGCCCATGGGTGCGATGGAGGCGGTGAGAGAGATAGAGAGGCTTCCCTCC 1339
DB TCCGTGGGGGAGCCCATGGGTGCGATGGAGGCGGTGAGAGAGATAGAGAGGCTTCCCTCC 600
QY 541 AAGAGGAGCCCGGGGTTTCCGAGGAGAAACCCCTCTGAGAGAGAAACCTCCAGCACCGAGC 600
DB AAGAGGAGCCCGGGGTTTCCGAGGAGAAACCCCTCTGAGAGAGAAACCTCCAGCACCGAGC 1399
QY 1340 AAGAGGAGCCCGGGGTTTCCGAGGAGAAACCCCTCTGAGAGAGAAACCTCCAGCACCGAGC 1399
DB AAGAGGAGCCCGGGGTTTCCGAGGAGAAACCCCTCTGAGAGAGAAACCTCCAGCACCGAGC 660
QY 601 TGAAGACTGGGAGTGTCCCAATCCGGGTTTGGAGAACCAAACTTCCCTGGAGAACAGA 660
DB TGAAGACTGGGAGTGTCCCAATCCGGGTTTGGAGAACCAAACTTCCCTGGAGAACAGA 1459
QY 1400 TGAAGACTGGGAGTGTCCCAATCCGGGTTTGGAGAACCAAACTTCCCTGGAGAACAGA 1459
DB TGAAGACTGGGAGTGTCCCAATCCGGGTTTGGAGAACCAAACTTCCCTGGAGAACAGA 720
QY 661 GTGCAACCAAGTGAAGGCGCCCAAGGCTTCCCTCCGCGCAACCTTCCGCGCCCC 720
DB GTGCAACCAAGTGAAGGCGCCCAAGGCTTCCCTCCGCGCAACCTTCCGCGCCCC 1519
QY 1460 GTGCAACCAAGTGAAGGCGCCCAAGGCTTCCCTCCGCGCAACCTTCCGCGCCCC 1519
DB GTGCAACCAAGTGAAGGCGCCCAAGGCTTCCCTCCGCGCAACCTTCCGCGCCCC 780
QY 721 GGGTGTGATCGTGGCAGAGGTGGCCCTGTGTGCAATGCGGAGAGAGAGAGTGGCTCAT 780
DB GGGTGTGATCGTGGCAGAGGTGGCCCTGTGTGCAATGCGGAGAGAGAGAGTGGCTCAT 1579
QY 1520 GGGTGTGATCGTGGCAGAGGTGGCCCTGTGTGCAATGCGGAGAGAGAGAGTGGCTCAT 1579
DB GGGTGTGATCGTGGCAGAGGTGGCCCTGTGTGCAATGCGGAGAGAGAGAGTGGCTCAT 840
QY 781 GGATCTGTGTGTCTCCGCTGGAATGTTCAAGGTGGCGCTGTGTGAGACAGAGTGGCTT 840
DB GGATCTGTGTGTCTCCGCTGGAATGTTCAAGGTGGCGCTGTGTGAGACAGAGTGGCTT 1639
QY 1580 GGATCTGTGTGTCTCCGCTGGAATGTTCAAGGTGGCGCTGTGTGAGACAGAGTGGCTT 1639
DB GGATCTGTGTGTCTCCGCTGGAATGTTCAAGGTGGCGCTGTGTGAGACAGAGTGGCTT 900
QY 841 CCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 900
DB CCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1639
QY 1640 CCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1639
DB CCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 960
QY 901 GGGGCCCCCTGGAACCTTTGATGGAACAGATGGGAGAGAGAGAGAGAGAGAGAGAGAGAG 960
DB GGGGCCCCCTGGAACCTTTGATGGAACAGATGGGAGAGAGAGAGAGAGAGAGAGAGAGAG 1759
QY 1700 GGGGCCCCCTGGAACCTTTGATGGAACAGATGGGAGAGAGAGAGAGAGAGAGAGAGAGAG 1759
DB GGGGCCCCCTGGAACCTTTGATGGAACAGATGGGAGAGAGAGAGAGAGAGAGAGAGAGAG
```

QY 961 TGAAGAAATGATTAAGGAGAGACACCTGAGAGGCGAGAGATCGGCTTAC 1012  
DB TGAAGAAATGATTAAGGAGAGACACCTGAGAGGCGAGAGATCGGCTTAC 1811

## RESULT 2

US-09-822-830A-49/c  
Sequence 49; Application US/09822830A  
Patent No. US20020142952A1  
GENERAL INFORMATION:  
APPLICANT: Genetics Institute, Inc.  
APPLICANT: Wong, Gordon G.  
APPLICANT: Clark, Hilary  
APPLICANT: Rechelet, Kim  
APPLICANT: Agostino, Michael J.  
APPLICANT: Howes, Steven H.  
APPLICANT: Resnick, Richard J.  
APPLICANT: Gulukota, Kamalakkar  
APPLICANT: Graham, James R.  
TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING NOVEL SECRETED PROTEINS  
FILE REFERENCE: GIN 6402  
CURRENT APPLICATION NUMBER: US/09/822,830A  
CURRENT FILING DATE: 2001-03-29  
PRIOR APPLICATION NUMBER: 60/195,604  
PRIOR FILING DATE: 2000-04-06  
NUMBER OF SEQ ID NOS: 631  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 49  
LENGTH: 2176  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-09-822-830A-49

Query Match 100.0%; Score 1012; DB 9; Length 2176;  
Best Local Similarity 100.0%; Pred. No. 1.5e-283;  
Matches 1012; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 AGGACGGGTGGAATGGGCGCTGGAGCGAGGTGGCTTCAATAGCTGTGGACC 60
DB AGGACGGGTGGAATGGGCGCTGGAGCGAGGTGGCTTCAATAGCTGTGGACC 1146
QY 1205 AGGACGGGTGGAATGGGCGCTGGAGCGAGGTGGCTTCAATAGCTGTGGACC 1146
DB AGGACGGGTGGAATGGGCGCTGGAGCGAGGTGGCTTCAATAGCTGTGGACC 1086
QY 61 CATGATGAAGACCAAGATCTTATCTAGGCCCCCTGTAGATCCAGATGAACCTCTGA 120
DB CATGATGAAGACCAAGATCTTATCTAGGCCCCCTGTAGATCCAGATGAACCTCTGA 1086
QY 1145 CATGATGAAGACCAAGATCTTATCTAGGCCCCCTGTAGATCCAGATGAACCTCTGA 1086
DB CATGATGAAGACCAAGATCTTATCTAGGCCCCCTGTAGATCCAGATGAACCTCTGA 180
QY 121 CAACAGTGAATTTATATACAGGATTTAAATGACAGTGTACTCTAGATGATCTGGCAGA 180
DB CAACAGTGAATTTATATACAGGATTTAAATGACAGTGTACTCTAGATGATCTGGCAGA 1026
QY 1085 CAACAGTGAATTTATATACAGGATTTAAATGACAGTGTACTCTAGATGATCTGGCAGA 1026
DB CAACAGTGAATTTATATACAGGATTTAAATGACAGTGTACTCTAGATGATCTGGCAGA 240
QY 181 CTCTTTAAGCAAGTGTGGGTTGTTAAGTGAACAGAGAACTGGGCAACCATGATCCA 240
DB CTCTTTAAGCAAGTGTGGGTTGTTAAGTGAACAGAGAACTGGGCAACCATGATCCA 966
QY 1025 CTCTTTAAGCAAGTGTGGGTTGTTAAGTGAACAGAGAACTGGGCAACCATGATCCA 966
DB CTCTTTAAGCAAGTGTGGGTTGTTAAGTGAACAGAGAACTGGGCAACCATGATCCA 300
QY 241 CATCTACCTGGAACAGAGAAACAGAGAAACCCAAAGGCGATGCCAGTGTCTATGAGA 300
DB CATCTACCTGGAACAGAGAAACAGAGAAACCCAAAGGCGATGCCAGTGTCTATGAGA 906
QY 965 CATCTACCTGGAACAGAGAAACAGAGAAACCCAAAGGCGATGCCAGTGTCTATGAGA 906
DB CATCTACCTGGAACAGAGAAACAGAGAAACCCAAAGGCGATGCCAGTGTCTATGAGA 360
QY 301 CCCACCACTGCAAGGCTGCGTGAATGTTGATGGGAAAGATTTCAAGGAGCA 360
DB CCCACCACTGCAAGGCTGCGTGAATGTTGATGGGAAAGATTTCAAGGAGCA 846
QY 905 CCCACCACTGCAAGGCTGCGTGAATGTTGATGGGAAAGATTTCAAGGAGCA 846
DB CCCACCACTGCAAGGCTGCGTGAATGTTGATGGGAAAGATTTCAAGGAGCA 420
QY 361 ACTTAAAGTCTCCCTTGTGCGAAGAACCTTCAATGAACATATGCGGGGTGTCTGCC 420
DB ACTTAAAGTCTCCCTTGTGCGAAGAACCTTCAATGAACATATGCGGGGTGTCTGCC 786
QY 845 ACTTAAAGTCTCCCTTGTGCGAAGAACCTTCAATGAACATATGCGGGGTGTCTGCC 786
DB ACTTAAAGTCTCCCTTGTGCGAAGAACCTTCAATGAACATATGCGGGGTGTCTGCC 480
QY 421 ACCCGTGAAGGAGAGGATGCAACCACTCCGTGAGAGTCCAGAGAGCCCGAGAGG 480
DB ACCCGTGAAGGAGAGGATGCAACCACTCCGTGAGAGTCCAGAGAGCCCGAGAGG 726
QY 785 ACCCGTGAAGGAGAGGATGCAACCACTCCGTGAGAGTCCAGAGAGCCCGAGAGG 726
DB ACCCGTGAAGGAGAGGATGCAACCACTCCGTGAGAGTCCAGAGAGCCCGAGAGG 540
QY 481 TCCGTGGGGGAGCCCATGGGTGCGATGGAGGCCGTGTGAGAGATGAGAGAGGCTTCCCTCC 540
DB TCCGTGGGGGAGCCCATGGGTGCGATGGAGGCCGTGTGAGAGATGAGAGAGGCTTCCCTCC 666
QY 725 TCCGTGGGGGAGCCCATGGGTGCGATGGAGGCCGTGTGAGAGATGAGAGAGGCTTCCCTCC 666
DB TCCGTGGGGGAGCCCATGGGTGCGATGGAGGCCGTGTGAGAGATGAGAGAGGCTTCCCTCC
```

QY 541 AGAGAGCCCCGGGGTTCGCGAGGAAACCCCTCTGAGGAGGAAAGCTCCAGCACCGAGC 600  
DB 665 AAGAGAGCCCCGGGGTTCGCGAGGAAACCCCTCTGAGGAGGAAAGCTCCAGCACCGAGC 606  
QY 601 TGGAGACTGGCAGTGTCCCAATCCGGGTGTGAAACAGAACTTCGCTGAGAAACA 660  
DB 605 TGGAGACTGGCAGTGTCCCAATCCGGGTGTGAAACAGAACTTCGCTGAGAAACA 546  
QY 661 GTGCAACAGTGTAAAGGCCCAAAAGCTGAAAGCTTCCTCCGACCTTCCTCCGCCCC 720  
DB 545 GTGCAACAGTGTAAAGGCCCAAAAGCTGAAAGCTTCCTCCGACCTTCCTCCGCCCC 486  
QY 721 GGGGTGTGATGTGAGGAGGTGGCCCTGTGATGCGGGGAGGAGAGGTGGCCCTCAT 780  
DB 485 GGGGTGTGATGTGAGGAGGTGGCCCTGTGATGCGGGGAGGAGAGGTGGCCCTCAT 426  
QY 781 GATCGTGTGTCTCCCGGTGAAATGTTCAAGGTGGCCGTGTGAGACAGAGGTGCTT 840  
DB 425 GATCGTGTGTCTCCCGGTGAAATGTTCAAGGTGGCCGTGTGAGACAGAGGTGCTT 366  
QY 841 CCGTGTGTGCGCGGGGACATGACCGAGGTGGCTTTGTGTGAGAAAGACGAGGTGGCTTG 900  
DB 365 CCGTGTGTGCGCGGGGACATGACCGAGGTGGCTTTGTGTGAGAAAGACGAGGTGGCTTG 306  
QY 901 GGGGCCCCCTGACCTTTGATGAAACAGATGGAGGAAAGAGAGAGAGAGAGAGAGAGAG 960  
DB 305 GGGGCCCCCTGACCTTTGATGAAACAGATGGAGGAAAGAGAGAGAGAGAGAGAGAGAG 246  
QY 961 TGGAAAAATGATTAAGAGCGAGCACCGTCAAGAGCGAGAGATCGGCTTAC 1012  
DB 245 TGGAAAAATGATTAAGAGCGAGCACCGTCAAGAGCGAGAGATCGGCTTAC 194

RESULT 3  
US-10-791-017A-1

Sequence 1, Application US/10791017A  
Publication No. US20040197827A1  
GENERAL INFORMATION:  
APPLICANT: JENAPHARM GmbH & Co. KG  
TITLE OF INVENTION: Methods for Determining Hormonal Effects of Substances  
FILE REFERENCE: Pac 3684/11  
CURRENT APPLICATION NUMBER: US/10/791.017A  
CURRENT FILING DATE: 2004-03-02  
NUMBER OF SEQ ID NOS: 7  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 1  
LENGTH: 2390  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (44)..(2011)  
OTHER INFORMATION: EMS  
US-10-791-017A-1

Query Match 100.0%; Score 1012; DB 18; Length 2390;  
Best Local Similarity 100.0%; Pred. No. 1.6e-283;  
Matches 1012; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 AGGAGCGGTGGAATGGGAGCGCTGGAGGCGAGGTGGCTTCAATTAAGCTGTGGACC 60  
DB 1000 AGGAGCGGTGGAATGGGAGCGCTGGAGGCGAGGTGGCTTCAATTAAGCTGTGGACC 1059  
QY 61 CATGATGAAGAGCAGATCTTATCTAGCCCACTGTAGATCCAGATGAAGACTCTGA 120  
DB 1060 CATGATGAAGAGCAGATCTTATCTAGCCCACTGTAGATCCAGATGAAGACTCTGA 1119  
QY 121 CAACAGTGCATTTATGTACAAGATTAATGACAGTGTACTCTAGATATCTGGACA 180  
DB 1120 CAACAGTGCATTTATGTACAAGATTAATGACAGTGTACTCTAGATATCTGGACA 1179  
QY 181 CTCTTTAAGCAGTGTGGGTGTGTTAAGTGAACAAGAGAACTGGGCAACCATGATCCA 240

DB 1180 CTCTTTAAGCAGTGTGGGTGTGTTAAGTGAACAAGAGAACTGGGCAACCATGATCCA 1239  
QY 241 CATCTACCTTGACAGAGAAACAGAAAGCCCAAGCGATGCCAAGTGTCTATGAGA 300  
DB 1240 CATCTACCTTGACAGAGAAACAGAAAGCCCAAGCGATGCCAAGTGTCTATGAGA 1299  
QY 301 CCCAACCACTGCCAAGCTGCGCGGTGAATGGTTTGAATGGGAAAGATTTTCAAGGAGCA 360  
DB 1300 CCCAACCACTGCCAAGCTGCGCGGTGAATGGTTTGAATGGGAAAGATTTTCAAGGAGCA 1359  
QY 361 ACTTAAAGTCTCCCTTGTCTGGAAGAGCTCCCAATGAACAGTATGCGGGGTGTGCC 420  
DB 1360 ACTTAAAGTCTCCCTTGTCTGGAAGAGCTCCCAATGAACAGTATGCGGGGTGTGCC 1419  
QY 421 ACCCGTGAAGGAGAGGAGATGCAACCACTCCGTGAGGTCCAGAGGCTCCAGAGG 480  
DB 1420 ACCCGTGAAGGAGAGGAGATGCAACCACTCCGTGAGGTCCAGAGGCTCCAGAGG 1479  
QY 481 TCGTGGGGGACCCATGGGTGCAATGAGGCGGTGGAGGAGATGAAGAGGCTTCCCTCC 540  
DB 1480 TCGTGGGGGACCCATGGGTGCAATGAGGCGGTGGAGGAGATGAAGAGGCTTCCCTCC 1539  
QY 541 AAGAGACCCCGGGGTTCCCGAGGAAACCCCTCTGAGAGAGAAACGTCCAGACCGAGC 600  
DB 1540 AAGAGACCCCGGGGTTCCCGAGGAAACCCCTCTGAGAGAGAAACGTCCAGACCGAGC 1599  
QY 601 TGGAGACTGGCAGTGTCCCAATCCGGTTGTGAAACAGAACTTGCTGAGAAACA 660  
DB 1600 TGGAGACTGGCAGTGTCCCAATCCGGTTGTGAAACAGAACTTGCTGAGAAACA 1659  
QY 661 GTGCAACAGTGTAAAGGCCCAAAAGCTTGAAGGCTTCTCCGACCTTCCTCCGCCCC 720  
DB 1660 GTGCAACAGTGTAAAGGCCCAAAAGCTTGAAGGCTTCTCCGACCTTCCTCCGCCCC 1719  
QY 721 GGGTGTGATCTGTGACAGAGGTGACCTGTGTGACATGCGGGGAGAAAGTGTGCTCAT 780  
DB 1720 GGGTGTGATCTGTGACAGAGGTGACCTGTGTGACATGCGGGGAGAAAGTGTGCTCAT 1779  
QY 781 GATCGTGTGTCTCCCGTGAATGTTCAAGGTGGCCGTGTGAGACAGAGGTGCTT 840  
DB 1780 GATCGTGTGTCTCCCGTGAATGTTCAAGGTGGCCGTGTGAGACAGAGGTGCTT 1839  
QY 841 CCGTGTGTGCGCGGGGACATGACCGAGGTGGCTTTGTGTGAGAAAGACGAGGTGGCTTG 900  
DB 1840 CCGTGTGTGCGCGGGGACATGACCGAGGTGGCTTTGTGTGAGAAAGACGAGGTGGCTTG 1899  
QY 901 GGGGCCCCCTGACCTTTGATGAAACAGATGGAGGAAAGAGAGAGAGAGAGAGAGAGAG 960  
DB 1900 GGGGCCCCCTGACCTTTGATGAAACAGATGGAGGAAAGAGAGAGAGAGAGAGAGAGAG 1959  
QY 961 TGGAAAAATGATTAAGAGCGAGCACCGTCAAGAGCGAGAGATCGGCTTAC 1012  
DB 1960 TGGAAAAATGATTAAGAGCGAGCACCGTCAAGAGCGAGAGATCGGCTTAC 2011

RESULT 4  
US-09-880-107-3769

Sequence 3769, Application US/09880107  
Patent No. US20020142981A1  
GENERAL INFORMATION:  
APPLICANT: Horne, Darci T.  
APPLICANT: Vockley, Joseph G.  
APPLICANT: Scherf, Uwe  
TITLE OF INVENTION: Gene Logic, Inc.  
FILE REFERENCE: 44921-5028-MO  
CURRENT APPLICATION NUMBER: US/09/880,107  
CURRENT FILING DATE: 2001-06-14  
PRIOR APPLICATION NUMBER: US 60/211,379  
PRIOR FILING DATE: 2000-06-14  
PRIOR APPLICATION NUMBER: US 60/237,054  
PRIOR FILING DATE: 2000-10-02  
NUMBER OF SEQ ID NOS: 3950

```

: SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO 3769
: LENGTH: 2350
: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE:
: OTHER INFORMATION: Genbank Accession No. US20020142981A1 X66899
US-03-880-107-3769

```

Query Match	99.8%	Score 1010.4;	DB 9;	Length 2390;
Best Local Similarity	99.9%	Pred. No. 4.6e-283;		
Matches 1011; Conservative	0;	Mismatches 1;	Indels 0;	Gaps 0;

OY	1	AGGACGCGGTGGAATGGGCAACGCGCTGGAGACGAGGTGGCTTCAATAGCCTGGTGGAC	60
Db	1000	AGGACGCGGTGGAATGGGCAACGCGCTGGAGACGAGGTGGCTTCAATAGCCTGGTGGAC	1059
OY			
OY	61	CATGATGTAAGACCAAGATCTTGAATCTTAGGCCCCACTGTAAATCCAGATGAAGCTCTGA	120
Db	1060	CATGATGTAAGACCAAGATCTTGAATCTTAGGCCCCCTGTGTAAATCCAGATGAAGCTCTGA	1119
OY	121	CAACGATGCAATTTATGTACAAAGATTAAATGACAGTGTGACTCTAGATGATCTGGCAGA	180
Db	1120	CAACGATGCAATTTATGTACAAAGATTAAATGACAGTGTGACTCTAGATGATCTGGCAGA	1179
OY			
OY	181	CTTCTTTAAGCAGTGTGGGTTTGTTAAGATGACAAAGAACTGGGCAACCCATGATCCA	240
Db	1180	CTTCTTTAAGCAGTGTGGGTTTGTTAAGATGACAAAGAACTGGGCAACCCATGATCCA	1239
OY	241	CATCTACCTGGACAAAGAAACAGGAAACCCAAAGGGGATGACACAGGTCTCTATGTAAGA	300
Db	1240	CATCTACCTGGACAAAGAAACAGGAAACCCAAAGGGATGACACAGGTCTCTATGTAAGA	1299
OY	301	CCCAACCACTGACAGAGCTGCGGTGGAATGGTTTGTATGGGAAGAATTTCAAGGAGCA	360
Db	1300	CCCAACCACTGACAGAGCTGCGGTGGAATGGTTTGTATGGGAAGAATTTCAAGGAGCA	1359
OY	361	ACTTAAAGTCTCCCTTGTCTGGGAAGAACCTTCCATGAAACATATGCGGGGTGGTCTGCC	420
Db	1360	ACTTAAAGTCTCCCTTGTCTGGGAAGAACCTTCCATGAAACATATGCGGGGTGGTCTGCC	1419
OY	421	ACCCGCTAGAGGACAGAGCATGCCACAACCTCCGTGAGAGGTCCAGAGGCCCCAGAGAG	480
Db	1420	ACCCGCTAGAGGACAGAGCATGCCACAACCTCCGTGAGAGGTCCAGAGGCCCCAGAGAG	1479
OY	481	TCCTGGGGAACCCATGAGGTCCGATGGGAAGCCGTGAGAGATAGAGAGGCTTCCCTCC	540
Db	1480	TCCTGGGGAACCCATGAGGTCCGATGGGAAGCCGTGAGAGATAGAGAGGCTTCCCTCC	1539
OY	541	AAGAGGACCCCGGGGGTTCCGAGGGGAACCCCTCTGGAAGAGAAAGTCCAGACCCAGAC	600
Db	1540	AAGAGGACCCCGGGGGTTCCGAGGGGAACCCCTCTGGAAGAGAAAGTCCAGACCCAGAC	1599
OY	601	TGGAGACTGGCAGTGTCCCAATCCGGGTTGTGAAACCAAGAATTGCTGCTGAGAAACGA	660
Db	1600	TGGAGACTGGCAGTGTCCCAATCCGGGTTGTGAAACCAAGAATTGCTGCTGAGAAACGA	1659
OY	661	GTGCAACCAAGTATAAGGCCCAACCTCGAAGGCTTCTCCGCCCAACCTTTCCGCCCCC	720
Db	1660	GTGCAACCAAGTATAAGGCCCAACCTCGAAGGCTTCTCTCCGCCCAACCTTTCCGCCCCC	1719
OY	721	GGGTGTGATCGTGGCAGAGTGTGGCCCTGTGTGACATGCGGGAGGAAGAGTGGCTCAT	780
Db	1720	GGGTGTGATCGTGGCAGAGTGTGGCCCTGTGTGACATGCGGGAGGAAGAGTGGCTCAT	1779
OY	781	GGATCTGTGTGTCCTCCGCTGGAATGTTCAAGCTGACCGTGTGAGACAAAGTGTGCTT	840
Db	1780	GGATCTGTGTGTCCTCCGCTGGAATGTTCAAGCTGACCGTGTGAGACAAAGTGTGCTT	1839
OY	841	CGGTGTGGCCGGGGACATGGAACGAGGTGGCTTTGGTGGAGGAACAAGAGTGGCCCTGG	900
Db	1840	CGGTGTGGCCGGGGACATGGAACGAGGTGGCTTTGGTGGAGGAACAAGAGTGGCCCTGG	1899

QY	901	GGGGCCCCCTGACCTTTGATGAAACATGAGGAAGAAAGAGAGACGTGAGAGACC	960
Db	1900	GGGGCCCCCTGACCTTTGATGAAACATGAGGAAGAAAGAGAGACGTGAGAGACC	195
QY	961	TGGAATAATGATTAAGCGCAGCACCGTCAAGAGCGCAGAGATCGSCCTTAC	1012
Db	1960	TGGAATAATGATTAAGCGCAGCACCGTCAAGAGCGCAGAGATCGSCCTTAC	2011

RESULT 5  
US-09-960-706-1081  
; Sequence 1081, Application US/09960706  
; Publication No. US20030134280A1

```

1  TITLE OF INVENTION: Identifying Drugs for and Diagnosis of Benign Prostatic Hyperplasia
2  TITLE OF INVENTION: Gene Expression Profiles
3  FILE REFERENCE: 44921-5029-01us
4  CURRENT APPLICATION NUMBER: US/09/960,706
5  CURRENT FILING DATE: 2001-09-24
6  PRIOR APPLICATION NUMBER: 60/223,323
7  PRIOR FILING DATE: 2000-08-07
8  PRIOR APPLICATION NUMBER: 09/873,319
9  PRIOR FILING DATE: 2001-06-05
10 NUMBER OF SEQ ID NOS: 1124
11 SOFTWARE: PatentIn Ver. 2.1
12 SEQ ID NO 1081
13 LENGTH: 2390
14 TYPE: DNA
15 ORGANISM: Homo sapiens
16 FEATURE:
17 OTHER INFORMATION: Genbank Accession No. US20030134280A1 X66899
18 US-09-960-706-1081

```

Query Match	Similarity	99.8%	Pred. No.	4.6e-283	DB 10	Length 2390
Best Local	Similarity	99.9%	Pred. No.	4.6e-283	DB 10	Length 2390
Matches 1011	Conservative	0	Mismatches	1	Indels	Gaps
QY	1	AGGACGGGCTGAGATATGGGACGCGCTGGAGAGCGAGGTGCTTCAATATAGCCGTGTGAC	60			
DB	1000	AGGACGGGCTGAGATATGGGACGCGCTGGAGAGCGAGGTGCTTCAATATAGCCGTGTGAC	1059			
QY	61	CATGATGAGAGACCAAGATCTTGATCTTAGGCCCACTGTAGATCCAGATGAAACTCTGA	120			
DB	1060	CATGATGAGAGACCAAGATCTTGATCTTAGGCCCTCTGTAGATCCAGATGAAACTCTGA	1119			
QY	121	CAAGAGTGCAATTTATATGACAAAGATTTAAATGACAGTGATCTCTAGATGATCGGCGA	180			
DB	1120	CAAGAGTGCAATTTATATGACAAAGATTTAAATGACAGTGATCTCTAGATGATCTCGGCGA	1179			
QY	181	CTTCTTTAAGCAGTGTGGGGTTGTTAAGATGAAACAAGAGAATTGGGCAACCCATGATCA	240			
DB	1180	CTTCTTTAAGCAGTGTGGGGTTGTTAAGATGAAACAAGAGAATTGGGCAACCCATGATCA	1239			
QY	241	CATCTACTGTGACAAAGAAACAGAAACCCAAAGCGCATGCCCAAGTGTCTATGTAGA	300			
DB	1240	CATCTACTGTGACAAAGAAACAGAAACCCAAAGCGCATGCCCAAGTGTCTATGTAGA	1299			
QY	301	CCCAACCACTGCCCAAGGCTGCGGTGAGATGTTGATGGGAAAGATTTTCAAGGACAA	360			
DB	1300	CCCAACCACTGCCCAAGGCTGCGGTGAGATGTTGATGGGAAAGATTTTCAAGGACAA	1359			
QY	361	ACTTAAAGTCTCCCTTGCTCGGAAGAAGCCTCCAAATGAACAGTATGCGGGGTGCTCTGC	420			
DB	1360	ACTTAAAGTCTCCCTTGCTCGGAAGAAGCCTCCAAATGAACAGTATGCGGGGTGCTCTGC	1419			
QY	421	ACCCGATGAGGACAGAGGATGCCAACCACTCCGTGAGGTCCAGAGGCCCAAGAGG	480			
DB	1420	ACCCGATGAGGACAGAGGATGCCAACCACTCCGTGAGGTCCAGAGGCCCAAGAGG	1479			
QY	481	TCTGTGGGGAACCCATGGGTGCGCATGGGAAGCGCTGTGAGAGAGATAGAGAGGCTTCCCTCC	540			
DB	1480	TCTGTGGGGAACCCATGGGTGCGCATGGGAAGCGCTGTGAGAGAGATAGAGAGGCTTCCCTCC	1539			



QY 541 AAGAGACCCCGGGGTTCCCGAGGAAACCCCTCTGAGAGAGAAAGTCCAGACCGAGC 600  
DB 1540 AAGAGACCCCGGGGTTCCCGAGGAAACCCCTCTGAGAGAGAAAGTCCAGACCGAGC 1599  
QY 601 TGGAGACTGGACAGTGTCCCAATCCGGGTGTGGAACCAAGACTTGCCTGAGAAACAGA 660  
DB 1600 TGGAGACTGGACAGTGTCCCAATCCGGGTGTGGAACCAAGACTTGCCTGAGAAACAGA 1659  
QY 661 GTGCAACCAAGTGAAGGCCCCCAAGGCTTCTCCGCAACCTTTCCGCCCC 720  
DB 1660 GTGCAACCAAGTGAAGGCCCCCAAGGCTTCTCCGCAACCTTTCCGCCCC 1719  
QY 721 GGGTGTGATCTGTGAGAGAGTGGCCCTGTGTGATCCGGGAGAGAAAGAGTGGCTCAT 780  
DB 1720 GGGTGTGATCTGTGAGAGAGTGGCCCTGTGTGATCCGGGAGAGAAAGAGTGGCTCAT 1779  
QY 781 GGATCGTGTGTCTCCGCTGGAATGTTCAAGAGTGGCCGTGTGAGAGACAGAGTGGCTT 840  
DB 1780 GGATCGTGTGTCTCCGCTGGAATGTTCAAGAGTGGCCGTGTGAGAGACAGAGTGGCTT 1839  
QY 841 CCGTGTGTGAGCGGGGCAATGACCGAGTGGCTTGTGTGAGAGAAACAGAGTGGCTTGG 900  
DB 1840 CCGTGTGTGAGCGGGGCAATGACCGAGTGGCTTGTGTGAGAGAAACAGAGTGGCTTGG 1899  
QY 901 GGGGCCCCCTTGACCTTTGATGAGACAGATGGAGAGAAAGAGAGAGAGTGGAGACC 960  
DB 1900 GGGGCCCCCTTGACCTTTGATGAGACAGATGGAGAGAAAGAGAGAGAGTGGAGACC 1959  
QY 961 TGGAAAAATGATTAAGGCGAGACCGTCAAGAGCGCAGAGATCGGCTTAC 1012  
DB 1960 TGGAAAAATGATTAAGGCGAGACCGTCAAGAGCGCAGAGATCGGCTTAC 2011

## RESULT 6

US-09-873-319-717  
; Sequence 717, Application US/09873319A  
; Publication No. US20030134324A1  
; GENERAL INFORMATION:  
; APPLICANT: Munger, William E.  
; APPLICANT: Kulkarni, Prakash  
; APPLICANT: Getzenberg, Robert H.  
; APPLICANT: Mega, Iwao  
; APPLICANT: Yamamoto, Jun  
; TITLE OF INVENTION: Identifying Drugs for and Diagnosis of Benign Prostatic  
; FILE REFERENCE: 44921-5029-US  
; CURRENT APPLICATION NUMBER: US/09/873,319A  
; EARLIER APPLICATION NUMBER: US 60/223,323  
; EARLIER FILING DATE: 2000-08-07  
; NUMBER OF SEQ ID NOS: 755  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 717  
; LENGTH: 2390  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; OTHER INFORMATION: Genbank Accession No. US20030134324A1 X66899  
US-09-873-319-717

Query Match 99.8%; Score 1010.4; DB 10; Length 2390;  
Best Local Similarity 99.9%; Pred. No. 4.6e-283;  
Matches 1011; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 AGGACGGGTGATGGGAGCGCTGAGAGCGAGGTGGTTCAATTAAGCTGTGGAGC 60  
DB 1000 AGGACGGGTGATGGGAGCGCGCTGAGAGCGAGGTGGTTCAATTAAGCTGTGGAGC 1059  
QY 61 CATGATGAAGACCAAGATCTTATCTAGGCCCACTGTAGATCCAGATGAAGACTCTGA 120  
DB 1060 CATGATGAAGACCAAGATCTTATCTAGGCCCTCTGTAGATCCAGATGAAGACTCTGA 1119

QY 121 CAACAGGCAATTTATGTAACAAGATTAAATGAACAGTGTACTCTAGATGATCGGAGCA 180  
DB 1120 CAACAGGCAATTTATGTAACAAGATTAAATGAACAGTGTACTCTAGATGATCGGAGCA 1179  
QY 181 CTTCTTTAAGCACTGTGGGTTGTAAAGATGAACAAGAACTGGGCAACCCATGATCCA 240  
DB 1180 CTTCTTTAAGCACTGTGGGTTGTAAAGATGAACAAGAACTGGGCAACCCATGATCCA 1239  
QY 241 CATCTACTGGAACAAGAAACAGAAAGCCCAAGGATCCACAGTGTCTTATGAAGA 300  
DB 1240 CATCTACTGGAACAAGAAACAGAAAGCCCAAGGATCCACAGTGTCTTATGAAGA 1299  
QY 301 CCCACCACTGCAAGGCTGGCGGTGAATGTGTTGAATGGGAAAGATTTTCAAGGAGCAA 360  
DB 1300 CCCACCACTGCAAGGCTGGCGGTGAATGTGTTGAATGGGAAAGATTTTCAAGGAGCAA 1359  
QY 361 ACTTAAAGTCTCCCTTGTCTGGAAGAGCTTCCAAATGAACAGTATCGGGGTGTCTGCC 420  
DB 1360 ACTTAAAGTCTCCCTTGTCTGGAAGAGCTTCCAAATGAACAGTATCGGGGTGTCTGCC 1419  
QY 421 ACCCGTGAAGGAGCAATGCAACCACTCCGTGAGGTCCAGAGAGCCCAAGAGG 480  
DB 1420 ACCCGTGAAGGAGCAATGCAACCACTCCGTGAGGTCCAGAGAGCCCAAGAGG 1479  
QY 481 TCCTGGGGGACCAATGGTTCGCAATGGAGGCGGTGAGAGAGATGAAGAGGCTTCCCTCC 540  
DB 1480 TCCTGGGGGACCAATGGTTCGCAATGGAGGCGGTGAGAGAGATGAAGAGGCTTCCCTCC 1539  
QY 541 AAGAGACCCCGGGGTTCCCGAGGAAACCCCTCTGAGAGAGAAAGTCCAGACCGAGC 600  
DB 1540 AAGAGACCCCGGGGTTCCCGAGGAAACCCCTCTGAGAGAGAAAGTCCAGACCGAGC 1599  
QY 601 TGGAGACTGGACAGTGTCCCAATCCGGGTGTGGAACCAAGACTTGCCTGAGAAACAGA 660  
DB 1600 TGGAGACTGGACAGTGTCCCAATCCGGGTGTGGAACCAAGACTTGCCTGAGAAACAGA 1659  
QY 661 GTGCAACCAAGTGAAGGCCCCCAAGGCTTCTCCGCAACCTTTCCGCCCC 720  
DB 1660 GTGCAACCAAGTGAAGGCCCCCAAGGCTTCTCCGCAACCTTTCCGCCCC 1719  
QY 721 GGGTGTGATCTGTGAGAGAGTGGCCCTGTGTGATCCGGGAGAGAAAGAGTGGCTTGG 780  
DB 1720 GGGTGTGATCTGTGAGAGAGTGGCCCTGTGTGATCCGGGAGAGAAAGAGTGGCTTGG 1779  
QY 781 GGATCGTGTGTCTCCGCTGGAATGTTCAAGAGTGGCCGTGTGAGAGACAGAGTGGCTT 840  
DB 1780 GGATCGTGTGTCTCCGCTGGAATGTTCAAGAGTGGCCGTGTGAGAGACAGAGTGGCTT 1839  
QY 841 CCGTGTGTGAGCGGGGCAATGACCGAGTGGCTTGTGTGAGAGAAACAGAGTGGCTTGG 900  
DB 1840 CCGTGTGTGAGCGGGGCAATGACCGAGTGGCTTGTGTGAGAGAAACAGAGTGGCTTGG 1899  
QY 901 GGGGCCCCCTTGACCTTTGATGAGACAGATGGAGAGAAAGAGAGAGAGTGGAGACC 960  
DB 1900 GGGGCCCCCTTGACCTTTGATGAGACAGATGGAGAGAAAGAGAGAGAGTGGAGACC 1959  
QY 961 TGGAAAAATGATTAAGGCGAGACCGTCAAGAGCGCAGAGATCGGCTTAC 1012  
DB 1960 TGGAAAAATGATTAAGGCGAGACCGTCAAGAGCGCAGAGATCGGCTTAC 2011

## RESULT 7

US-09-822-830A-410/c  
; Sequence 410, Application US/09822830A  
; Patent No. US20020142952A1  
; GENERAL INFORMATION:  
; APPLICANT: Genetics Institute, Inc.  
; APPLICANT: Wong, Gordon G.  
; APPLICANT: Clark, Hilary  
; APPLICANT: Fechtel, Kim  
; APPLICANT: Agostino, Michael J.  
; APPLICANT: Howes, Steven H.  
; APPLICANT: Resnick, Richard J.

APPLICANT: Gulukota, Kamalakar  
APPLICANT: Graham, James R.  
FILE OF INVENTION: POLYNUCLEOTIDES ENCODING NOVEL SECRETED PROTEINS  
FILE REFERENCE: GIN 6402  
CURRENT APPLICATION NUMBER: US/09/822,830A  
CURRENT FILING DATE: 2001-03-29  
PRIOR APPLICATION NUMBER: 60/195,604  
PRIOR FILING DATE: 2000-04-06  
NUMBER OF SEQ ID NOS: 631  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 410  
LENGTH: 2273  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-09-822-830A-410

Query Match 87.2%; Score 882.4; DB 9; Length 2273;  
Best Local Similarity 94.6%; Pred. No. 7.5e-246;  
Matches 957; Conservative 0; Mismatches 1; Indels 54; Gaps 2;

QY 1 AGAAGCGGTGGAATGAGGAGCGCTGAGAGCGAGGTGCTTCAATTAAGCTGTGAGCC 60  
DB 1310 AGAAGCGGTGGAATGAGGAGCGCTGAGAGCGAGGTGCTTCAATTAAGCTGTGAGCC 1254  
QY 61 CATGATGAGAGCAGACATCTGATCTAGGCGCCACTGTAGATCCAGATGAAGCTCTGA 120  
DB 1253 CATGATGAGAGCAGACATCTGATCTAGGCGCCACTGTAGATCCAGATGAAGCTCTGA 1194  
QY 121 CAACAGTGAATTTATGTACAAAGATTAAATGACAGTGTACTTATAGATCTGTGAGCA 180  
DB 1193 CAACAGTGAATTTATGTACAAAGATTAAATGACAGTGTACTTATAGATCTGTGAGCA 1134  
QY 181 CTCTTTAAAGCAGTGGGCTTGTATAGATGAACAAGAACTGGGCAACCCATGATCCA 240  
DB 1133 CTCTTTAAAGCAGTGGGCTTGTATAGATGAACAAGAACTGGGCAACCCATGATCCA 1074  
QY 241 CATCTACCTGGAACAAGAAACAAGAAAGCCAAAGGCGATGCCAAGTGTCTATGAGA 300  
DB 1073 CATCTACCTGGAACAAGAAACAAGAAAGCCAAAGGCGATGCCAAGTGTCTATGAGA 1014  
QY 301 CCCAACCCTGCGCAAGGCTGCGGTGAAATGTTGATGAGAAAGATTTCAAGGAGCAA 360  
DB 1013 CCCAACCCTGCGCAAGGCTGCGGTGAAATGTTGATGAGAAAGATTTCAAGGAGCAA 954  
QY 361 ACTTTAAAGTCTCCCTGCTGCTGGAAGAAAGCTTCCAAATGAAGTATGCGGGTGTCTGCC 420  
DB 953 ACTTTAAAGTCTCCCTGCTGCTGGAAGAAAGCTTCCAAATGAAGTATGCGGGTGTCTGCC 894  
QY 421 ACCCGGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 480  
DB 893 ACCCGGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 834  
QY 481 TCTTGGGGGAGCCCAATGGGTGCAATGGAGGCGGTGAGAGATGAGAGAGGCTTCCCTCC 540  
DB 833 TCTTGGGGGAGCCCAATGGGTGCAATGGAGGCGGTGAGAGATGAGAGAGGCTTCCCTCC 774  
QY 541 AAGAGAGCCCCGGGGTCTCCGAGAGGAAACCCCTCTGAGAGAGAAACCTCCAGACAGAGC 600  
DB 773 AAGAGAGCCCCGGGGTCTCCGAGAGGAAACCCCTCTGAGAGAGAAACCTCCAGACAGAGC 714  
QY 601 TGGAGACTGGCAGTGTCCCAATCCGGGTGTGAGAAACAGAACTTCCCTGAGAGACAGA 660  
DB 713 TGGAGACTGGCAGTGTCCCAATCCGGGTGTGAGAAACAGAACTTCCCTGAGAGACAGA 654  
QY 661 GTGCAACCAAGTGAAGGCCCAAAAGCTGAAGGCTTCTCCCGCACCTTTCCGCCCC 720  
DB 653 GTGCAACCAAGT----- 643  
QY 721 GGGGTGTGATCTGTGAGAGAGTGGCCCTGTGAGATCCGGGAGAGAAAGATGGCTCAT 780  
DB 642 --GTGGTGTGATCTGTGAGAGAGTGGCCCTGTGAGATCCGGGAGAGAAAGATGGCTCAT 585  
QY 781 GGATCTGTGTGTCTCCGGTGAATGTTCAAGAGTGTGCGGTGTGAGACAGAGAGTGGCTT 840

DB 584 GGATCTGTGTGTCTCCGGTGAATGTTCAAGAGTGGCCCTGTGAGAGAGATGGCTT 525  
QY 841 CCGTGTGCGCGGGGATGAGACCGAGAGTGGCTTTGTGAGAGAAACAGAGTGGCTTGG 900  
DB 524 CCGTGTGCGCGGGGATGAGACCGAGAGTGGCTTTGTGAGAGAAACAGAGTGGCTTGG 465  
QY 901 GGGGCCCCCTGAGACCTTGTATGGAACAGATGGAGAGAGAAAGAGAGAGCGTGGAGAGC 960  
DB 464 GGGGCCCCCTGAGACCTTGTATGGAACAGATGGAGAGAGAAAGAGAGAGCGTGGAGAGC 405  
QY 961 TGGAAATATGATTAAGGAGAGACCGTCAAGAGCGCAGAGATGGCCCTTAC 1012  
DB 404 TGGAAATATGATTAAGGAGAGACCGTCAAGAGCGCAGAGATGGCCCTTAC 353

RESULT 8  
US-10-198-846-9847  
Sequence 9847, Application US/10198846  
Publication No. US2003009974A1  
GENERAL INFORMATION:  
APPLICANT: Lillie, James  
APPLICANT: Xu, Yongyao  
APPLICANT: Wang, Youzhen  
APPLICANT: Steilmann, Kathleen  
TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS  
TITLE OF INVENTION: FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND  
THERAPY OF BREAST CANCER  
FILE REFERENCE: MRI-049  
CURRENT APPLICATION NUMBER: US/10/198,846  
CURRENT FILING DATE: 2002-07-18  
PRIOR APPLICATION NUMBER: 60/306,220  
PRIOR FILING DATE: 2001-07-18  
NUMBER OF SEQ ID NOS: 14084  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 9847  
LENGTH: 2299  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-10-198-846-9847

Query Match 56.3%; Score 570; DB 14; Length 2299;  
Best Local Similarity 100.0%; Pred. No. 5.3e-155;  
Matches 570; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 56 GGAACCATGATGAGAGAGACCAATCTTGTATAGGCTCCACCTGTATATCAATGAAGAC 115  
DB 1072 GGAACCATGATGAGAGAGACCAATCTTGTATAGGCTCCACCTGTATATCAATGAAGAC 1131  
QY 116 TCTGACACAGTGCATTTATGATGACAAAGATTAAATGACAGTGTGACTAGATGATCTG 175  
DB 1132 TCTGACACAGTGCATTTATGATGACAAAGATTAAATGACAGTGTGACTAGATGATCTG 1191  
QY 176 GCAGACTTCTTAAAGCAGTGGGCTTGTAAAGATGAACAAGAACTGGGCAACCATG 225  
DB 1192 GCAGACTTCTTAAAGCAGTGGGCTTGTAAAGATGAACAAGAACTGGGCAACCATG 1251  
QY 236 ATCCATATCTACTTGAACAAGAAACAAGAAAGCCAAAGGCGATGCTCAAGTGTCTAT 295  
DB 1252 ATCCATATCTACTTGAACAAGAAACAAGAAAGCCAAAGGCGATGCTCAAGTGTCTAT 1211  
QY 296 GAGAGCCACCCCACTGCAAGGCTGCGGTGAGATGGTTGATGGAGAAAGATTTCAAGG 355  
DB 1312 GAGAGCCACCCCACTGCAAGGCTGCGGTGAGATGGTTGATGGAGAAAGATTTCAAGG 1371  
QY 356 AGCAACTTAAGTCTCCCTGTCTCGAAGAAAGCTTCAATGAACAGTATGCGGGTGTG 415  
DB 1372 AGCAACTTAAGTCTCCCTGTCTCGAAGAAAGCTTCAATGAACAGTATGCGGGTGTG 1431  
QY 416 CTGCAACCCCGTGAAGGAGGAGGAGGATGCCACCACTCCGTGTGAGAGTCCAGAGAGGCCA 475  
DB 1432 CTGCAACCCCGTGAAGGAGGAGGAGGATGCCACCACTCCGTGTGAGAGTCCAGAGAGGCCA 1491

Qy 476 GGAAGTCTCTGGGGGACCCATGGTGCATGGGAGGCGGTGGAGGAGATAGAGAGGCTTC 535  
Db 1492 GGAAGTCTCTGGGGGACCCATGGTGCATGGGAGGCGGTGGAGGAGATAGAGAGGCTTC 1551  
Qy 536 CTTCCAGAGAGACCCCGGGGTTCCCGAGGAGACCCCTCTGGAGGAGAGAAAGTCCAGACAC 595  
Db 1552 CTTCCAGAGAGACCCCGGGGTTCCCGAGGAGACCCCTCTGGAGGAGAGAAAGTCCAGACAC 1611  
Qy 596 CGAGCTGGAGACTGGCATGTGTCCTCAATCCG 625  
Db 1612 CGAGCTGGAGACTGGCATGTGTCCTCAATCCG 1641

## RESULT 9

US-09-918-995-27690  
; Sequence 27690, Application US/09918995  
; Publication No. US20030073623A1  
; GENERAL INFORMATION:  
; APPLICANT: Hvaeq, Inc.  
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED  
; FILE REFERENCE: 20411-756  
; CURRENT APPLICATION NUMBER: US/09/918,995  
; PRIOR FILING DATE: 2001-07-30  
; PRIOR APPLICATION NUMBER: US/09/235,076  
; NUMBER OF SEQ ID NOS: 38054  
; SOFTWARE: FASTSEQ for Windows Version 3.0  
; SEQ ID NO 27690  
; LENGTH: 550  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)...(550)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-918-995-27690

Query Match 50.2%; Score 508.4; DB 10; Length 550;  
Best Local Similarity 97.5%; Pred. No. 2.6e-137;  
Matches 515; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

Qy 335 GATGGGAAGATTTTCAAGGAGCAAACTTAATCTCCCTGTCGGAAGAAAGCTTCCA 394  
Db 22 GAAGAGGCAATTTCTTCAAGGAGCAAACTTAATCTCCCTGTCGGAAGAAAGCTTCCA 81  
Qy 395 ATGAACAGTATGCGGGGTGTGCTGCAACCCGTCGAGGAGGAGCATGCCACCACTC 454  
Db 82 ATGAACAGTATGCGGGGTGTGCTGCAACCCGTCGAGGAGGAGCATGCCACCACTC 141  
Qy 455 CGTGAAGTCCAGAGGCGCCAGAGGTCTTGAGGGAACCATGGTGCATGGAGGCGT 514  
Db 142 CGTGAAGTCCAGAGGCGCCAGAGGTCTTGAGGGAACCATGGTGCATGGAGGCGT 201  
Qy 515 GGAAGGATGAGAGGCTTCCCTCCAGAGAGACCCCGGGGTTCCCGAGGAGAAAGCTTCT 574  
Db 202 GGAAGGATGAGAGGCTTCCCTCCAGAGAGACCCCGGGGTTCCCGAGGAGAAAGCTTCT 261  
Qy 575 GGAAGGAGAAAGCTCCAGCAACGAGCTGAGACTGGAGTGTCCCAATCCGGGTTGGGA 634  
Db 262 GGAAGGAGAAAGCTCCAGCAACGAGCTGAGACTGGAGTGTCCCAATCCGGGTTGGGA 321  
Qy 635 AACCAAGACTTCCCTGAGAAAGAGTGCACCAAGTGAAGGCCCAAGGCTTGAAGGC 694  
Db 322 AACCAAGACTTCCCTGAGAAAGAGTGCACCAAGTGAAGGCCCAAGGCTTGAAGGC 381  
Qy 695 TTCTCTCCCGCACCCTTCCGCGCCCGGGGTGTATGTGCAAGAGTGGCCCTGTGGGC 754  
Db 382 TTCTCTCCCGCACCCTTCCGCGCCCGGGGTGTATGTGCAAGAGTGGCCCTGTGGGC 441  
Qy 755 ATCCGGGAGAGAGAGTGCCTCATGATCTGTGTGTCCTCCGGTGAAGTTCAGAGT 814  
Db 442 ATCCGGGAGAGAGAGTGCCTCATGATCTGTGTGTCCTCCGGTGAAGTTCAGAGT 501

Qy 815 GGCCTGTGTGAGACAGAGGTGCTTCCGTGTGCGCCGGGAGATGAC 862  
Db 502 GGCCTGTGTGAGACAGAGGTGCTTCCGTGTGCGCCGGGAGATGAC 549

## RESULT 10

US-09-918-995-9557  
; Sequence 9557, Application US/09918995  
; Publication No. US20030073623A1  
; GENERAL INFORMATION:  
; APPLICANT: Hvaeq, Inc.  
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED  
; FILE REFERENCE: 20411-756  
; CURRENT APPLICATION NUMBER: US/09/918,995  
; PRIOR FILING DATE: 2001-07-30  
; PRIOR APPLICATION NUMBER: US/09/235,076  
; NUMBER OF SEQ ID NOS: 38054  
; SOFTWARE: FASTSEQ for Windows Version 3.0  
; SEQ ID NO 9557  
; LENGTH: 568  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)...(568)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-918-995-9557

Query Match 43.2%; Score 437.2; DB 10; Length 568;  
Best Local Similarity 97.5%; Pred. No. 1.3e-116;  
Matches 466; Conservative 0; Mismatches 8; Indels 4; Gaps 2;

Qy 1 AGAAGCGGTGGAATGAGGAGCGCTGGAGAGGAGGCTTCAATAGCTGTGAGC 60  
Db 94 AGAAGCGGTGGAATGAGGAGCGCTGGAGAGGAGGCTTCAATAGCTGTGAGC 150  
Qy 61 CATGATGAGAGACAGATCTTATGATGAGGCGCCAGCTGTAGATCCAGATGAGACTGGA 120  
Db 151 CATGATGAGAGACAGATCTTATGATGAGGCGCCAGCTGTAGATCCAGATGAGACTGGA 210  
Qy 121 CAACAGTCAATTTATGATCAAGATTAATGAAGTGTGATCTGTAGATATCTGGAGA 180  
Db 211 CAACAGTCAATTTATGATCAAGATTAATGAAGTGTGATCTGTAGATATCTGGAGA 270  
Qy 181 CTCTTTAAGCATGTGGGTTGTTAAGATGAACAAGAACTGGGCAACCATGATCCA 240  
Db 271 CTCTTTAAGCATGTGGGTTGTTAAGATGAACAAGAACTGGGCAACCATGATCCA 330  
Qy 241 CATCTACTGAGCAAGAAACAGAAAGCCCAAGGCGATGCCAGATGTCTTATGAAGA 300  
Db 331 CATCTACTGAGCAAGAAACAGAAAGCCCAAGGCGATGCCAGATGTCTTATGAAGA 390  
Qy 301 CCCACCACTGCGAAGGCTGCGGTGAATGTTGATGGAAAGATTTTCAAGGAGCA 360  
Db 391 CCCACCACTGCGAAGGCTGCGGTGAATGTTGATGGAAAGATTTTCAAGGAGCA 450  
Qy 361 ACTTAAAGTCTCCCTGCTGGAAGAAAGCTCCAAAGAAAGATGCGGGGTGTCTGCC 420  
Db 451 ACTTAAAGTCTCCCTGCTGGAAGAAAGCTCCAAAGAAAGATGCGGGGTGTCTGCC 509  
Qy 421 ACCCGTGAAGGAGAGAGTGCACCACTCCGTGAGAGTCCAGAGAGCCAGGA 478  
Db 510 ACCCGTGAAGGAGAGAGTGCACCACTCCGTGAGAGTCCAGAGAGCCAGGA 567

## RESULT 11

US-09-925-301-669  
; Sequence 669, Application US/09925301  
; Patent No. US20020052308A1  
; GENERAL INFORMATION:

```

APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
FILE REFERENCE: P106
CURRENT APPLICATION NUMBER: US/09/925,301
CURRENT FILING DATE: 2001-08-10
PRIOR APPLICATION NUMBER: PCT/US00/05882
PRIOR FILING DATE: 2000-03-08
PRIOR APPLICATION NUMBER: 60/124,270
PRIOR FILING DATE: 1999-03-12
NUMBER OF SEQ ID NOS: 1694
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 669
LENGTH: 545
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
LOCATION: (8)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc_feature
LOCATION: (11)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc_feature
LOCATION: (13)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc_feature
LOCATION: (58)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc_feature
LOCATION: (337)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc_feature
LOCATION: (453)
OTHER INFORMATION: n equals a,t,g, or c
US-09-925-301-669

```

Query Match 30.5%; Score 309; DB 9; Length 545;  
Best Local Similarity 78.9%; Pred. No. 2,5e-79;

Matches 396; Conservative 0; Mismatches 93; Indels 13; Gaps 2;

```

467 GGAGGCCCAAGAGTCTCTGGGGGACCCATGGTGCATGGAGGCCCTGAGGAGATAGA 526
57 GNCGGCGCTCTAGATAGATGATCCCGGCTGAGGAAATTCGACAGAGATAGA 116
527 GGAGGCTTCCCTCAAGAGAGACCCCGGGTTCCCGAGGAAACCCCTCGAGAGAGAAAC 586
117 GGAGGCTTCCCTCAAGAGAGACCCCGGGTTCCCGAGGAAACCCCTCGAGAGAGAAAC 176
587 GTCCAGACCGAGCTGAGACTGCACTGTCCTCAATCCGGTGTGGAAACCAAACTTC 646
177 GTCCAGACCGAGCTGAGACTGCACTGTCCTCAATCCGGTGTGGAAACCAAACTTC 224
647 GCCTGAGAACAGAGTGCAACCACTGTAAGGCCCAAGCCTGAAGCCTTCCTCCGCCA 706
225 GATTTCTGTGTATGTAATGTAATGATGAGGGGTTGGAAACCAAACTTCCTGAGAGA 284
707 CCGTTCCGCGCCCGGGTGTGATCTGTGGAGAGAGTGGCCCTGTGCAAGCGGGAGAGA 766
285 ACAGAGTGCAACCACTGTGTGATCTGTGGAGAGAGTGGCCCTGTGCAAGCGGGAGAGA 344
767 AGAGTGAGCTCATGATCTGTGTGATCTGTGGAGAGAGTGGCCCTGTGCAAGCGGGAGAGA 826
345 AGAGTGAGCTCATGATCTGTGTGATCTGTGGAGAGAGTGGCCCTGTGCAAGCGGGAGAGA 404
827 GACAGAGTGAGCTTCTGTGTGATCTGTGGAGAGAGTGGCCCTGTGCAAGCGGGAGAGA 886
405 GACAGAGTGAGCTTCTGTGTGATCTGTGGAGAGAGTGGCCCTGTGCAAGCGGGAGAGA 464
887 CGAGTGAGCTTCTGTGTGATCTGTGGAGAGAGTGGCCCTGTGCAAGCGGGAGAGA 946
465 CGAGTGAGCTTCTGTGTGATCTGTGGAGAGAGTGGCCCTGTGCAAGCGGGAGAGA 523
947 GAGTGAGAGAGCTGAGAAAGA 968

```

Db 524 GGAGCTGAGAGAGCTGGAGAAA 545

RESULT 12

US-09-960-352-3851

Sequence 3851, Application US/09960352

Patent No. US20020137139A1

GENERAL INFORMATION:

APPLICANT: Warren, Wesley C.

APPLICANT: Tao, Nengbing

APPLICANT: Byatt, John C.

APPLICANT: Mathalagan, Nagappan

TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND

FILE REFERENCE: 16511.006/37-21(10298)C

CURRENT APPLICATION NUMBER: US/09/960,352

CURRENT FILING DATE: 2001-09-24

NUMBER OF SEQ ID NOS: 15112

SEQ ID NO 3851

LENGTH: 361

TYPE: DNA

ORGANISM: Bos taurus

OTHER INFORMATION: Clone ID: 17-LIB3057-024-Q1-K1-E1

US-09-960-352-3851

Query Match 26.7%; Score 269.8; DB 9; Length 361;

Best Local Similarity 84.2%; Pred. No. 5.5e-68;

Matches 304; Conservative 0; Mismatches 57; Indels 0; Gaps 0;

```

197 GGAGTGTGAAGTGAACAAGAGAACTGGCAACCCATGATCCATCTAATCTGACAAAG 256
1 GGAGTGTGAAGTGAACAAGAGAACTGGCAACCCATGATCCATCTAATCTGACAAAG 60
257 GAAACGGAAGAGCCCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 316
61 TAAACGGAAGAGAGCCCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 120
317 GCTGCGGTGAAGTGTGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 376
121 GCTGCGGTGAAGTGTGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 180
377 GCTGCGGTGAAGTGTGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 436
181 GCTGCGGTGAAGTGTGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 240
437 GGCATGCCACACCACTTCCGTGAGAGTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAG 496
241 GGCATGCCACACCACTTCCGTGAGAGTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAG 300
497 GGTGCGATGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 556
301 GGTGCGATGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360
557 T 557
361 T 361

```

RESULT 13

US-09-864-761-8728

Sequence 8728, Application US/09864761

Patent No. US20020048763A1

GENERAL INFORMATION:

APPLICANT: Penn, Sharon G.

APPLICANT: Rank, David R.

APPLICANT: Hanzel, David K.

APPLICANT: Chen, Wensheng

TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR

FILE REFERENCE: Aeomica-X-1

CURRENT APPLICATION NUMBER: US/09/864,761

CURRENT FILING DATE: 2001-05-23



```
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 18163
; LENGTH: 262
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO CHR22.1
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2.3
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 3.1
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 4.6
; OTHER INFORMATION: SWISSPROT HIT: Q07050, EVALUE 2.00e+00
; OTHER INFORMATION: NT HIT: X73003.1, EVALUE 0.00e+00
; OTHER INFORMATION: EST_HUMAN HIT: BB97260.1, EVALUE 0.00e+00
; US-10-029-386-18163

Query Match      25.1%; Score 254.2; DB 16; Length 262;
Best Local Similarity 98.8%; Pred. No. 1,7e-63;
Matches 256; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 717 CCCCCGGTGTGATCGTGGCAAGCTGCCCTGTGTCATGCGGGAGAGAGGTGGCC 776
    |||||||
Db 262 CCTTAGGTGTGATCGTGGCAAGCTGCCCTGTGTCATGCGGGAGAGAGGTGGCC 203

QY 777 TCATGATCGTGTGTCGCCGTGGATGTTCAAGGTGGCCGTGTGAGACAGAGTG 836
    |||||||
Db 202 TCATGATCGTGTGTCGCCGTGGATGTTCAAGGTGGCCGTGTGAGACAGAGTG 143

QY 837 GCTTCCGTGTGTCGCCGGCATGACCAGGTGCTTTGGTGAAGAGACGAGTGGCC 896
    |||||||
Db 142 GCTTCCGTGTGTCGCCGGCATGACCAGGTGCTTTGGTGAAGAGAGAGTGGCC 83

QY 897 CTGGGGGGGCCCTCTGACCTTTGATGGAACAGATGGAGAGAGAGAGAGTGGAG 956
    |||||||
Db 82 CTGGGGGGGCCCTCTGACCTTTGATGGAACAGATGGAGAGAGAGAGAGAGTGGAG 23

QY 957 GACCTGGAATAATGATTA 975
    |||||||
Db 22 GACCTGGAATAATGATTA 4
```

Search completed: February 20, 2005, 22:47:58  
Job time : 534.573 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 20, 2005, 04:03:23 ; Search time 2989.19 Seconds  
(without alignments)  
12886.762 Million cell updates/sec

Title: US-10-791-017a-1\_COPY\_1000\_2011

Perfect score: 1012  
Sequence: 1 aggacgcggtggaatgggca.....aggcagagatcgccctac 1012

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*  
1: gb\_est1:\*  
2: gb\_est2:\*  
3: gb\_est3:\*  
4: gb\_est4:\*  
5: gb\_est5:\*  
6: gb\_est6:\*  
7: gb\_est7:\*  
8: gb\_est8:\*  
9: gb\_est9:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1012	100.0	2073	3	CR593724	CR593724 full-length
2	1012	100.0	2083	3	CR608666	CR608666 full-length
3	1012	100.0	2103	3	CR619493	CR619493 full-length
4	1012	100.0	2157	3	CR604952	CR604952 full-length
5	1012	100.0	2314	3	CR608375	CR608375 full-length
6	996	98.4	2070	3	CR602561	CR602561 full-length
7	996	98.4	2093	3	CR594868	CR594868 full-length
8	996	98.4	2126	3	CR610888	CR610888 full-length
9	996	98.4	2159	3	CR625247	CR625247 full-length
10	892	88.1	2236	3	CR604639	CR604639 full-length
11	882.6	87.2	1379	3	AK014366	AK014366 Mus muscu
12	879.2	86.9	2107	3	AK019460	AK019460 Mus muscu
13	876	86.6	2373	3	AK034755	AK034755 Mus muscu
14	834.8	82.5	1020	5	BX387336	BX387336 BX387336
15	792.8	78.3	2269	5	AK049743	AK049743 Mus muscu
16	792.2	78.3	895	5	BX408406	BX408406 BX408406
17	789.6	78.0	1057	5	BX353312	BX353312 BX353312
18	789.4	78.0	929	5	BQ881305	BQ881305 AGENCOURT
19	784	77.5	1068	5	BX449773	BX449773 BX449773
20	778	76.9	1048	5	BX381313	BX381313 BX381313
21	777	76.8	962	5	BX407870	BX407870 BX407870
22	770.6	76.1	1013	5	BX465099	BX465099 BX465099
23	769	76.0	1028	5	BX399069	BX399069 BX399069
24	754.4	74.5	838	4	BG574437	BG574437 602596384

25	753.4	74.4	778	4	BG756215	BG756215 602713525
26	751	74.2	1064	5	BX417171	BX417171 BX417171
27	748.6	74.0	1062	5	BX443495	BX443495 BX443495
28	745.6	73.7	1055	5	BX428311	BX428311 BX428311
29	741.2	73.2	792	5	BX368129	BX368129 BX368129
30	740.6	73.2	933	5	BX387305	BX387305 BX387305
31	739.6	73.1	1032	5	BX407825	BX407825 BX407825
32	737.4	72.9	925	5	BX339942	BX339942 BX339942
33	733.8	72.5	964	5	BX382192	BX382192 BX382192
34	732.6	72.4	837	4	BG751179	BG751179 602729679
35	732.4	72.4	1038	5	BX335119	BX335119 BX335119
36	730.2	72.2	930	5	BX347606	BX347606 BX347606
37	722.6	71.7	901	2	BE562296	BE562296 601344870
38	715.4	70.7	787	2	BE741903	BE741903 601594075
39	712.4	70.4	928	5	BX346753	BX346753 BX346753
40	711.4	70.3	897	6	BY713354	BY713354 BY713354
41	701.4	69.3	726	4	BG754333	BG754333 602709891
42	699.6	69.1	951	5	BQ642355	BQ642355 AGENCOURT
43	697	68.9	849	2	BE563744	BE563744 601335129
44	694.6	68.6	843	7	CN162640	CN162640 952042 MA
45	691.8	68.4	805	5	BX449612	BX449612 BX449612

## ALIGNMENTS

RESULT 1  
LOCUS CR593724  
DEFINITION full-length cDNA clone CS0D1011Y005 of Placenta Cot 25-normalized  
of Homo sapiens (human).  
ACCESSION CR593724.1 GI:50474531  
VERSION CR593724.1  
KEYWORDS HTC; CNS/IT CDNA.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1. (bases 1 to 2073)  
AUTHORS Li, W.B., Gruber, C., Jessee, D., and Polayes, D.  
TITLE Full-length cDNA libraries and normalization  
JOURNAL Unpublished  
REMARK Contact: Feng Liang Email: fliang@life.techn.com URL: http://fulllength.invitrogen.com/ Invitrogen Corporation 1600 Faraday Avenue  
2 (bases 1 to 2073)  
REFERENCE Genoscope.  
AUTHORS Direct Submission  
JOURNAL Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail: seqref@genoscope.cns.fr - Web: www.genoscope.cns.fr)  
COMMENT 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies, a division of Invitrogen.  
FEATURES  
source  
1. 2073  
/organism="Homo sapiens"  
/db\_type="mRNA"  
/cd\_xref="taxon:9606"  
/clone="CS0D1011Y005"  
/issue\_type="Placenta Cot 25-normalized"  
/plasmid="pCMVSPORT\_6"  
ORIGIN  
Query Match 100.0%; Score 1012; DB 3; Length 2073;  
Best Local Similarity 100.0%; Pred. No. 8e-241;  
Matches 1012; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Oy 1 AGGACCGGTGGAATGGAGCGCTGAGAGCGAGGTTGAATTAAGCTGTGGAC 60  
Db 908 AGGACCGGTGGAATGGAGCGCTGAGAGCGAGGTTGAATTAAGCTGTGGAC 967

QY 61 CATGATGAAAGACAGATCTTGTATCTAGGCCCACTGTATGATCCAGATGAAAGACTCTGA 120  
 DB 968 CATGATGAAAGACAGATCTTGTATCTAGGCCCACTGTATGATCCAGATGAAAGACTCTGA 1027  
 QY 121 CAACAGTGAATTTATGATCAAGATTTAATGACAGTGTACTCTAGATGATCTGACAGA 180  
 DB 1028 CAACAGTGAATTTATGATCAAGATTTAATGACAGTGTACTCTAGATGATCTGACAGA 1087  
 QY 181 CTTCTTTAAGCAGTGTGGGTTGTAAAGATGAAACAAGAACTGGGCAACCATGATCCA 240  
 DB 1088 CTTCTTTAAGCAGTGTGGGTTGTAAAGATGAAACAAGAACTGGGCAACCATGATCCA 1147  
 QY 241 CATCTACTGGAACAAGAAACAAGAAAGCCCAAGGATGCCACAGTGTCTATGAAAGA 300  
 DB 1148 CATCTACTGGAACAAGAAACAAGAAAGCCCAAGGATGCCACAGTGTCTATGAAAGA 1207  
 QY 301 CCCAACCACTGCAAGAGCTGCGTGGAAATGTTGATGGGAAAGATTTTCAAGGAGCAA 360  
 DB 1208 CCCAACCACTGCAAGAGCTGCGTGGAAATGTTGATGGGAAAGATTTTCAAGGAGCAA 1267  
 QY 361 ACTTAAAGTCTCCCTTGTCTGGAAGAAAGCTCCATGAAACATATGCGGGGTGCTGCC 420  
 DB 1268 ACTTAAAGTCTCCCTTGTCTGGAAGAAAGCTCCATGAAACATATGCGGGGTGCTGCC 1327  
 QY 421 ACCCGGTGAGGGCAGAGGCAATGCCACCACTCCGTGGAGTTCAGAGAGCCCAAGAGG 480  
 DB 1328 ACCCGGTGAGGGCAGAGGCAATGCCACCACTCCGTGGAGTTCAGAGAGCCCAAGAGG 1387  
 QY 481 TCCGTGGGGGACCCATGAGGTGCGATGGAGGCGGTGAGAGAGATPAGAGAGGCTTCCCTCC 540  
 DB 1388 TCCGTGGGGGACCCATGAGGTGCGATGGAGGCGGTGAGAGAGATPAGAGAGGCTTCCCTCC 1447  
 QY 541 AAGAGAACCCCGGGGTTCCGAGAGGAAACCCCTCTGAGAGAGAAACGTCCAGACCGAGC 600  
 DB 1448 AAGAGAACCCCGGGGTTCCGAGAGGAAACCCCTCTGAGAGAGAAACGTCCAGACCGAGC 1507  
 QY 601 TGGAGATGAGCAATGTCCCAATCCGGGTGTGGAACAACAATCTCCCTGAGAGAAAGA 660  
 DB 1508 TGGAGATGAGCAATGTCCCAATCCGGGTGTGGAACAACAATCTCCCTGAGAGAAAGA 1567  
 QY 661 GTGCAACAGTGTAAAGGCCCAAGAGCTTGAAGGCTTCTCCCGCACCTTTCCGCCCC 720  
 DB 1568 GTGCAACAGTGTAAAGGCCCAAGAGCTTGAAGGCTTCTCCCGCACCTTTCCGCCCC 1627  
 QY 721 GGGTGTGATCTGTGCAAGAGGTGAGCCCTGTGTGAGTCCGGGAGAAAGAGTGGCTCAT 780  
 DB 1628 GGGTGTGATCTGTGCAAGAGGTGAGCCCTGTGTGAGTCCGGGAGAAAGAGTGGCTCAT 1687  
 QY 781 GGAATCGTGTGTGTCCTCCGTGGAATGTTCAAGAGTGGCCGTGTGGAACAAGAGTGGCTT 840  
 DB 1688 GGAATCGTGTGTGTCCTCCGTGGAATGTTCAAGAGTGGCCGTGTGGAACAAGAGTGGCTT 1747  
 QY 841 CCGTGTGTGAGCGGGGCAATGACCGAGGTGCTTGTGTGAGAGAGAGAGAGTGGCTTGG 900  
 DB 1748 CCGTGTGTGAGCGGGGCAATGACCGAGGTGCTTGTGTGAGAGAGAGAGAGTGGCTTGG 1807  
 QY 901 GGGGCCCCCTGTGACCTTTGATGAAACAAGTGGAGAAAGAGAGAGAGAGAGAGAGAGAG 960  
 DB 1808 GGGGCCCCCTGTGACCTTTGATGAAACAAGTGGAGAAAGAGAGAGAGAGAGAGAGAGAG 1867  
 QY 961 TGGAAAAATGATTAAGCGAGCACTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1012  
 DB 1868 TGGAAAAATGATTAAGCGAGCACTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1919

RESULT 2  
 LOCUS CR608666 2083 bp mRNA linear HTC 21-JUL-2004  
 DEFINITION full-length cDNA clone CS0DG004YL23 of B cells (Ramos cell line) of  
 Homo sapiens (human).  
 ACCESSION CR608666  
 VERSION CR608666.1 GI:50489473

KEYWORDS HTC; CNSLT cDNA.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 2083)  
 AUTHORS Li, W.B., Gruber, C., Jessee, J., and Polayres, D.  
 TITLE Full-length cDNA libraries and normalization  
 JOURNAL Unpublished  
 REMARK Contact : Feng Liang Email : fliang@life.techn.com URL :  
 http://fulllength.invitrogen.com/ Invitrogen Corporation 1600  
 Faraday Avenue  
 Redwood City, CA 94063  
 REFERENCE 2 (bases 1 to 2083)  
 AUTHORS Direct Submission  
 TITLE Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage :  
 JOURNAL BP 191 91006 Evry cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)  
 COMMENT - Web : www.genoscope.cns.fr  
 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime  
 end enriched, double-strand cDNA was digested with Not I and cloned  
 into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library  
 was normalized. Library was constructed by Life Technologies, a  
 division of Invitrogen.  
 FEATURES  
 source  
 1..2083  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="CS0DG004YL23"  
 /issue\_type="B cells (Ramos cell line)"  
 /plasmid="pCMVSPORT\_6"  
 ORIGIN  
 Query Match 100.0%; Score 1012; DB 3; Length 2083;  
 Best Local Similarity 100.0%; Pred. No. 8.1e-241;  
 Matches 1012; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGGAGCGGGTGGAAATGAGGAGCGCTGAGAGAGAGTGGCTTCAATAGGCTGTGAGAC 60  
 DB 908 AGGAGCGGGTGGAAATGAGGAGCGCTGAGAGAGAGTGGCTTCAATAGGCTGTGAGAC 967  
 QY 61 CATGATGAAAGACAGATCTTGTATCTAGGCCCACTGTATGATCCAGATGAAAGACTCTGA 120  
 DB 968 CATGATGAAAGACAGATCTTGTATCTAGGCCCACTGTATGATCCAGATGAAAGACTCTGA 1027  
 QY 121 CAACAGTGAATTTATGATCAAGATTTAATGACAGTGTACTCTAGATGATCTGACAGA 180  
 DB 1028 CAACAGTGAATTTATGATCAAGATTTAATGACAGTGTACTCTAGATGATCTGACAGA 1087  
 QY 181 CTTCTTTAAGCAGTGTGGGTTGTAAAGATGAAACAAGAACTGGGCAACCATGATCCA 240  
 DB 1088 CTTCTTTAAGCAGTGTGGGTTGTAAAGATGAAACAAGAACTGGGCAACCATGATCCA 1147  
 QY 241 CATCTACTGGAACAAGAAACAAGAAAGCCCAAGGATGCCACAGTGTCTATGAAAGA 300  
 DB 1148 CATCTACTGGAACAAGAAACAAGAAAGCCCAAGGATGCCACAGTGTCTATGAAAGA 1207  
 QY 301 CCCAACCACTGCAAGAGCTGCGTGGAAATGTTGATGGGAAAGATTTTCAAGGAGCAA 360  
 DB 1208 CCCAACCACTGCAAGAGCTGCGTGGAAATGTTGATGGGAAAGATTTTCAAGGAGCAA 1267  
 QY 361 ACTTAAAGTCTCCCTTGTCTGGAAGAAAGCTCCATGAAACATATGCGGGGTGCTGCC 420  
 DB 1268 ACTTAAAGTCTCCCTTGTCTGGAAGAAAGCTCCATGAAACATATGCGGGGTGCTGCC 1327  
 QY 421 ACCCGGTGAGGGCAGAGGCAATGCCACCACTCCGTGGAGTTCAGAGAGCCCAAGAGG 480  
 DB 1328 ACCCGGTGAGGGCAGAGGCAATGCCACCACTCCGTGGAGTTCAGAGAGCCCAAGAGG 1387  
 QY 481 TCCGTGGGGGACCCATGAGGTGCGATGGAGGCGGTGAGAGAGATPAGAGAGGCTTCCCTCC 540  
 DB 1388 TCCGTGGGGGACCCATGAGGTGCGATGGAGGCGGTGAGAGAGATPAGAGAGGCTTCCCTCC 1447

Qy 541 AAGAGACCCCGGGGTTCCCGAGGAGAACCCCTCTGAGAGAGAAAGCTCCAGCAGCAGC 600  
| | | | |  
Db 1448 AAGAGACCCCGGGGTTCCCGAGGAGAACCCCTCTGAGAGAGAAAGCTCCAGCAGCAGC 1507  
| | | | |  
Qy 601 TGGAGACTGCACTGTCCCAATCCGAGGTGTGAGAAACAGAACTTCCCTGAGAGACAGA 660  
| | | | |  
Db 1508 TGGAGACTGCACTGTCCCAATCCGAGGTGTGAGAAACAGAACTTCCCTGAGAGACAGA 1567  
| | | | |  
Qy 661 GTGCAACCACTGTAAAGGCCCCAAGCCTGAAAGGCTTCTCCGCCACCTTTCCGCCCC 720  
| | | | |  
Db 1568 GTGCAACCACTGTAAAGGCCCCAAGCCTGAAAGGCTTCTCCGCCACCTTTCCGCCCC 1627  
| | | | |  
Qy 721 GGGTGGTATCGTGGCAGAGGTGGCCTGTGGCATCGGGGAGAGAGAGTGGCTCAT 780  
| | | | |  
Db 1628 GGGTGGTATCGTGGCAGAGGTGGCCTGTGGCATCGGGGAGAGAGAGTGGCTCAT 1687  
| | | | |  
Qy 781 GATCGTGTGTCTCCCGGTGAAATGTTCAAGAGTGGCCGTGTGAGACAGAGTGGCTT 840  
| | | | |  
Db 1688 GATCGTGTGTCTCCCGGTGAAATGTTCAAGAGTGGCCGTGTGAGACAGAGTGGCTT 1747  
| | | | |  
Qy 841 CCGTGTGTGCGCGGGGCATGACCCGAGGTGGCTTTGTGTGAGAGAAACAGAGTGGCTT 900  
| | | | |  
Db 1748 CCGTGTGTGCGCGGGGCATGACCCGAGGTGGCTTTGTGTGAGAGAAACAGAGTGGCTT 1807  
| | | | |  
Qy 901 GGGGCCCCCTGACCTTTGATGAAACAGATGGAGAGAAAGAGAGAGTGGAGAC 960  
| | | | |  
Db 1808 GGGGCCCCCTGACCTTTGATGAAACAGATGGAGAGAAAGAGAGAGTGGAGAC 1867  
| | | | |  
Qy 961 TGGAAAAATGATTAAGGCGAGACCCGTCAAGAGCGAGAGATCGGCTTAC 1012  
| | | | |  
Db 1868 TGGAAAAATGATTAAGGCGAGACCCGTCAAGAGCGAGAGATCGGCTTAC 1919  
| | | | |

RESULT 3  
CR619493 2103 bp mRNA 1linear HTC 21-JUL-2004  
LOCUS CR619493  
DEFINITION full-length cDNA clone CSOD1076Y021 of Placenta Cot 25-normalized  
of Homo sapiens (human).  
ACCESSION CR619493  
VERSION CR619493.1 GI:50500300  
KEYWORDS HTC; CNS/IT cDNA.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens

REFERENCE  
AUTHORS Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
JOURNAL Li, W.-B., Gruber, C., Jessee, J., and Polayes, D.  
TITLE 1 (bases 1 to 2103)  
Unpublished  
Contact : Feng Liang Email : fliang@lifetech.com URL :  
http://fulllength.invitrogen.com/Invitrogen Corporation 1600  
REMARK Faraday Avenue  
Genoscope.  
2 (bases 1 to 2103)  
Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage :  
BP 191 91006 EVRY cedex - FRANCE (E-mail : seque@genoscope.cns.fr)  
- Web : www.genoscope.cns.fr  
1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime  
into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library  
was normalized. Library was constructed by Life Technologies, a  
division of Invitrogen.

## FEATURES

source  
1. .2103  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="CSOD1076Y021"  
/issue\_type="Placenta Cot 25-normalized"  
/plasmid="pCMVSPORT\_6"

ORIGIN

Query Match 100.0%; Score 1012; DB 3; Length 2103;  
Best Local Similarity 100.0%; Pred. No. 8.1e-241; Indels 0; Gaps 0;  
Matches 1012; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 AGAAGCGGTGGAATGAGGAGCGCTGAGAGAGAGAGTGGCTTCAATTAAGCTGGTGAAC 60  
| | | | |  
Db 945 AGAAGCGGTGGAATGAGGAGCGCTGAGAGAGAGAGTGGCTTCAATTAAGCTGGTGAAC 1004  
| | | | |  
Qy 61 CATGATGAAGAGACCAAGATCTTATCTAGGCCCACTGTAGATCCAGATGAAGCTTGA 120  
| | | | |  
Db 1005 CATGATGAAGAGACCAAGATCTTATCTAGGCCCACTGTAGATCCAGATGAAGCTTGA 1064  
| | | | |  
Qy 121 CAACAGTGCATTTATGTAAGATTAATGACAGTGTACTGATGATGATCTGGAGAGA 180  
| | | | |  
Db 1065 CAACAGTGCATTTATGTAAGATTAATGACAGTGTACTGATGATGATCTGGAGAGA 1124  
| | | | |  
Qy 181 CTTCTTTAAGCAGTGTGGGTGTTGTTAAGATGAACAAGAACTGGGCAACCATGATCCA 240  
| | | | |  
Db 1125 CTTCTTTAAGCAGTGTGGGTGTTGTTAAGATGAACAAGAACTGGGCAACCATGATCCA 1184  
| | | | |  
Qy 241 CATCTACCTTGAACAGAAACAGAAAGCCCAAGGATGCCACAGTGTCTATGAGA 300  
| | | | |  
Db 1185 CATCTACCTTGAACAGAAACAGAAAGCCCAAGGATGCCACAGTGTCTATGAGA 1244  
| | | | |  
Qy 301 CCCAACCACCTGCCAAGGCTGCGGTGAAATGTTGATGAGAAAGATTTTCAAGGAGCAA 360  
| | | | |  
Db 1245 CCCAACCACCTGCCAAGGCTGCGGTGAAATGTTGATGAGAAAGATTTTCAAGGAGCAA 1304  
| | | | |  
Qy 361 ACTTAAAGTCTCCCTTGTCTGGAAGAAAGCTCCCAATGAACAGTATGCGGGGTGCTGCC 420  
| | | | |  
Db 1305 ACTTAAAGTCTCCCTTGTCTGGAAGAAAGCTCCCAATGAACAGTATGCGGGGTGCTGCC 1364  
| | | | |  
Qy 421 ACCCTGTGAGGAGAGAGGATGCTCACCACTCTGTGAGGTCCAGAGAGCCCAAGAG 480  
| | | | |  
Db 1365 ACCCTGTGAGGAGAGAGGATGCTCACCACTCTGTGAGGTCCAGAGAGCCCAAGAG 1424  
| | | | |  
Qy 481 TCTTGGGGGACCCATGAGGTGTCATGAGAGGCGCTGAGAGAGATGAGAGAGCTTCCCTCC 540  
| | | | |  
Db 1425 TCTTGGGGGACCCATGAGGTGTCATGAGAGGCGCTGAGAGAGATGAGAGAGCTTCCCTCC 1484  
| | | | |  
Qy 541 AAGAGACCCCGGGGTTCCCGAGGAGAACCCCTCTGAGAGAGAAACGTCCAGCACCGAGC 600  
| | | | |  
Db 1485 AAGAGACCCCGGGGTTCCCGAGGAGAACCCCTCTGAGAGAGAAACGTCCAGCACCGAGC 1544  
| | | | |  
Qy 601 TGGAGACTGCACTGTCCCAATCCGAGGTGTGAGAAACAGAACTTGCCTGAGAAACAGA 660  
| | | | |  
Db 1545 TGGAGACTGCACTGTCCCAATCCGAGGTGTGAGAAACAGAACTTGCCTGAGAAACAGA 1604  
| | | | |  
Qy 661 GTGCAACCACTGTAAAGGCCCCAAGCCTGAAAGGCTTCTCCGCCACCTTTCCGCCCC 720  
| | | | |  
Db 1605 GTGCAACCACTGTAAAGGCCCCAAGCCTGAAAGGCTTCTCCGCCACCTTTCCGCCCC 1664  
| | | | |  
Qy 721 GGGTGGTATCGTGGCAGAGGTGGCCTGTGTGATCGGGGAGAGAGAGTGGCTCAT 780  
| | | | |  
Db 1665 GGGTGGTATCGTGGCAGAGGTGGCCTGTGTGATCGGGGAGAGAGAGTGGCTCAT 1724  
| | | | |  
Qy 781 GATCGTGTGTCTCCCGGTGAAATGTTCAAGAGTGGCCGTGTGAGACAGAGTGGCTT 840  
| | | | |  
Db 1725 GATCGTGTGTCTCCCGGTGAAATGTTCAAGAGTGGCCGTGTGAGACAGAGTGGCTT 1784  
| | | | |  
Qy 841 CCGTGTGTGCGCGGGGCATGACCCGAGGTGGCTTTGTGTGAGAGAAACAGAGTGGCTT 900  
| | | | |  
Db 1785 CCGTGTGTGCGCGGGGCATGACCCGAGGTGGCTTTGTGTGAGAGAAACAGAGTGGCTT 1844  
| | | | |  
Qy 901 GGGGCCCCCTGACCTTTGATGAAACAGATGGAGAGAAAGAGAGAGTGGAGAGC 960  
| | | | |  
Db 1845 GGGGCCCCCTGACCTTTGATGAAACAGATGGAGAGAAAGAGAGAGTGGAGAGC 1904  
| | | | |  
Qy 961 TGGAAAAATGATTAAGGCGAGACCCGTCAAGAGCGAGAGATCGGCTTAC 1012  
| | | | |  
Db 1905 TGGAAAAATGATTAAGGCGAGACCCGTCAAGAGCGAGAGATCGGCTTAC 1956  
| | | | |

RESULT 4  
LOCUS CR604952 2157 bp mRNA linear HTC 21-JUL-2004  
DEFINITION full-length cDNA clone CS0D1010Y112 of Placenta Cot 25-normalized  
of Homo sapiens (human).  
ACCESSION CR604952  
VERSION CR604952.1 GI:50485759  
KEYWORDS HTC; CNSLT\_CDNA.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 2157)  
AUTHORS Li W.B., Gruber C., Jessee J. and Polyes D.  
TITLE Full-length cDNA libraries and normalization  
JOURNAL Unpublished  
REMARK Contact: Feng Liang Email: fliang@lifetech.com URL: http://fulllength.invitrogen.com/Invitrogen Corporation 1600 Faraday Avenue  
2 (bases 1 to 2157)  
REFERENCE Genoscope.  
AUTHORS Direct Submission  
TITLE Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage :  
JOURNAL BP 191 91006 EVRY cedex - FRANCE (E-mail: seqref@genoscope.cns.fr  
- Web : www.genoscope.cns.fr)  
COMMENT 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime  
end enriched, double-strand cDNA was digested with Not I and cloned  
into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library  
was normalized. Library was constructed by Life Technologies, a  
division of Invitrogen.  
FEATURES  
source location/Qualifiers  
1..2157  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="CS0D1010Y112"  
/issue\_type="Placenta Cot 25-normalized"  
/plasmid="pCMVSPORT\_6"

ORIGIN  
Query Match 100.0%; Score 1012; DB 3; Length 2157;  
Best Local Similarity 100.0%; Pred. No. 8.1e-241;  
Matches 1012; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 AGAAGCGGCTGGATGGGACGCTGGAGACGAGGCTTCAATAGCTGTGAC 60  
DB 997 AGGACGGGCTGGATGGGACGCTGGAGACGAGGCTTCAATAGCTGTGAC 1056  
QY 61 CATGATGAAGAGCAAGATCTTGTATAGGCCCACTGTAGATCAGATGAAGACTG 120  
DB 1057 CATGATGAAGAGCAAGATCTTGTATAGGCCCACTGTAGATCAGATGAAGACTG 1116  
QY 121 CAACAGTGCAATTTATGTACAAAGATTAAATGACAGTGTGACTGTAGATCTGG 180  
DB 1117 CAACAGTGCAATTTATGTACAAAGATTAAATGACAGTGTGACTGTAGATCTGG 1176  
QY 181 CTTCTTTAAGAGGTGGGCTTTTAAGATGAACAAGAACTGGGCAACCCATGATCCA 240  
DB 1177 CTTCTTTAAGAGGTGGGCTTTTAAGATGAACAAGAACTGGGCAACCCATGATCCA 1236  
QY 241 CATCTACCTGGACAAGAAACAGAAAGCCCAAGAGCGATGCCAGATGTCCTATGAGA 300  
DB 1237 CATCTACCTGGACAAGAAACAGAAAGCCCAAGAGCGATGCCAGATGTCCTATGAGA 1296  
QY 301 CCCAACCACCTGCAAGGCTGCGTGGATGGTTTGTATGGGAAGATTTCAGGGAGCAA 360  
DB 1297 CCCAACCACCTGCAAGGCTGCGTGGATGGTTTGTATGGGAAGATTTCAGGGAGCAA 1356  
QY 361 ACTTAAAGTCTCCCTGTGCTGGAAAGAGCTTCAATGAACGATATGGGGGTGTCTGCC 420  
DB 1357 ACTTAAAGTCTCCCTGTGCTGGAAAGAGCTTCAATGAACGATATGGGGGTGTCTGCC 1416  
QY 421 ACCCGTAGAGGAGGAGGATGCAACCACTCCGTGAGAGTCCAGAGGCCAGAGAG 480

DB 1417 ACCCGTAGAGGAGGAGGATGCAACCACTCCGTGAGAGTCCAGAGGCCAGAGAG 1476  
QY 481 TCTTGGGGGAGCCCAATGGGTGCTGCAATGGAGAGCCGTGAGAGAGTATAGAGAGCTTCCCTCC 540  
DB 1477 TCTTGGGGGAGCCCAATGGGTGCTGCAATGGAGAGCCGTGAGAGAGTATAGAGAGCTTCCCTCC 1536  
QY 541 AAGAGAGCCCGGGGGTTCGAGAGGAGACCCCTCTGAGAGAGAAAGTCCAGACCGAGC 600  
DB 1537 AAGAGAGCCCGGGGGTTCGAGAGGAGACCCCTCTGAGAGAGAAAGTCCAGACCGAGC 1596  
QY 601 TGGAGACTGCAAGTGTCCCAATCCGGGTGTGAAAACAGAACTTGCTGAGAAACAG 660  
DB 1597 TGGAGACTGCAAGTGTCCCAATCCGGGTGTGAAAACAGAACTTGCTGAGAAACAG 1656  
QY 661 GTGCAACCAAGTGTAAAGCCCAAGCCCTTAAGGCTTCCCGGCCACCTTCCGGCCCC 720  
DB 1657 GTGCAACCAAGTGTAAAGCCCAAGCCCTTAAGGCTTCCCGGCCACCTTCCGGCCCC 1716  
QY 721 GGGTGTGATCGTGGCAGAGGTGACCTGTGATGCGGGAGAGAGAGTGGCTCAT 780  
DB 1717 GGGTGTGATCGTGGCAGAGGTGACCTGTGATGCGGGAGAGAGAGTGGCTCAT 1776  
QY 781 GGATCGTGTGTGTCCTGCGTGAATGTTCAAGAGTGGCCGTGTGAGACAGAGTGGCTT 840  
DB 1777 GGATCGTGTGTGTCCTGCGTGAATGTTCAAGAGTGGCCGTGTGAGACAGAGTGGCTT 1836  
QY 841 CCGTGTGTCGGGGGATGAGACGAGAGTGGCTTGTGTGAGAGAAACGAGTGGCTTGG 900  
DB 1837 CCGTGTGTCGGGGGATGAGACGAGAGTGGCTTGTGTGAGAGAAACGAGTGGCTTGG 1896  
QY 901 GGGGCCCCCTGACCTTGTATGGAACAGATGGAGAGAGAGAGAGAGTGGAGAGC 960  
DB 1897 GGGGCCCCCTGACCTTGTATGGAACAGATGGAGAGAGAGAGAGAGTGGAGAGC 1956  
QY 961 TGGAAAATGATTAAGAGGAGACCGTCAAGAGCGCAGAGATCGGCCCTTAC 1012  
DB 1957 TGGAAAATGATTAAGAGGAGACCGTCAAGAGCGCAGAGATCGGCCCTTAC 2008

RESULT 5  
LOCUS CR608375 2314 bp mRNA linear HTC 21-JUL-2004  
DEFINITION full-length cDNA clone CS0D1039YJ24 of Placenta Cot 25-normalized  
of Homo sapiens (human).  
ACCESSION CR608375  
VERSION CR608375.1 GI:50489182  
KEYWORDS HTC; CNSLT\_CDNA.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 2314)  
AUTHORS Li W.B., Gruber C., Jessee J. and Polyes D.  
TITLE Full-length cDNA libraries and normalization  
JOURNAL Unpublished  
REMARK Contact: Feng Liang Email: fliang@lifetech.com URL: http://fulllength.invitrogen.com/Invitrogen Corporation 1600 Faraday Avenue  
2 (bases 1 to 2314)  
REFERENCE Genoscope.  
AUTHORS Direct Submission  
TITLE Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage :  
JOURNAL BP 191 91006 EVRY cedex - FRANCE (E-mail: seqref@genoscope.cns.fr  
- Web : www.genoscope.cns.fr)  
COMMENT 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime  
end enriched, double-strand cDNA was digested with Not I and cloned  
into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library  
was normalized. Library was constructed by Life Technologies, a  
division of Invitrogen.  
FEATURES  
source location/Qualifiers  
1..2314  
/organism="Homo sapiens"

## ORIGIN

/mol\_type="mRNA"  
/db\_xref="taxon.9606"  
/clone="CS01039J24"  
/issue\_type="Placenta Cot 25-normalized"  
/plasmid="PCMVSPORT\_6"

Query Match 100.0%; Score 1012; DB 3; Length 2314;

Best Local Similarity 100.0%; Pred. No. 8.2e-241; Mismatches 0; Indels 0; Gaps 0;

Matches 1012; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 AGGACCGCGTGAATGGGACCGCTGGAGAGCGAGTGGCTTCAATAGCCTGGTGGACC 60
DB 997 AGGACCGCGTGAATGGGACCGCTGGAGAGCGAGTGGCTTCAATAGCCTGGTGGACC 1056
QY 61 CATGATGAAGACACAGATCTTGAATGAAGCCACTGTAGATCAGATGAAGACTCTGA 120
DB 1057 CATGATGAAGACACAGATCTTGAATGAAGCCACTGTAGATCAGATGAAGACTCTGA 1116
QY 121 CAACAGTGCATTTATATACAGATTTAATGACAGTGTGACTTATATATGATCTGGCAGA 180
DB 1117 CAACAGTGCATTTATATACAGATTTAATGACAGTGTGACTTATATATGATCTGGCAGA 1176
QY 181 CTCTTTAAGCAGTGTGGGTTGTATAGATGAACAAGAACTGGGACCACTGATCCA 240
DB 1177 CTCTTTAAGCAGTGTGGGTTGTATAGATGAACAAGAACTGGGACCACTGATCCA 1236
QY 241 CATCTAAGTGAACAGAGAAACAGAAAGCCCAAGCGATGCCAAGTGTCTATGAAGA 300
DB 1237 CATCTAAGTGAACAGAGAAACAGAAAGCCCAAGCGATGCCAAGTGTCTATGAAGA 1296
QY 301 CCCACCACTGCGCAAGGCTGCGTGGATGATGTTGATGGAAAGATTTTCAAGGAGCAA 360
DB 1297 CCCACCACTGCGCAAGGCTGCGTGGATGATGTTGATGGAAAGATTTTCAAGGAGCAA 1356
QY 361 ACTTAAAGTCTCCCTTGTCTGGAAGAGCCCTCAATGAAGATGAGGAGGCTGCGC 420
DB 1357 ACTTAAAGTCTCCCTTGTCTGGAAGAGCCCTCAATGAAGATGAGGAGGCTGCGC 1416
QY 421 ACCCGGTGAGGAGAGAGGATGCGACCACTCCCTGAGAGTCCAGAGAGCCGAGAGG 480
DB 1417 ACCCGGTGAGGAGAGAGGATGCGACCACTCCCTGAGAGTCCAGAGAGCCGAGAGG 1476
QY 481 TCTTGGGAGACCAATGGGTGCAATGGAGGCGTGGAGAGATGAGAGGCTTCCCTCC 540
DB 1477 TCTTGGGAGACCAATGGGTGCAATGGAGGCGTGGAGAGATGAGAGGCTTCCCTCC 1536
QY 541 AAGAGGACCCGCGGGTCCCGAGAGGAAACCCCTGAGAGAGAACTCCAGACCGAGC 600
DB 1537 AAGAGGACCCGCGGGTCCCGAGAGGAAACCCCTGAGAGAGAACTCCAGACCGAGC 1596
QY 601 TGGAGACTGCAATGTCATCCAGATCCGAGTGTGGAACACAGAACTTCCCTGAGAGCAGA 660
DB 1597 TGGAGACTGCAATGTCATCCAGATCCGAGTGTGGAACACAGAACTTCCCTGAGAGCAGA 1656
QY 661 GTGCAACCAAGTGAAGGCCCAAAAGCTGAAGGCTTCTCCCGCACCTTTCCGCCCC 720
DB 1657 GTGCAACCAAGTGAAGGCCCAAAAGCTGAAGGCTTCTCCCGCACCTTTCCGCCCC 1716
QY 721 GGGTGGTGAATCGTGGAGAGAGTGGCCCTGTGTGCAATGCGGGAGAGAAAGAGTGGCTCAT 780
DB 1717 GGGTGGTGAATCGTGGAGAGAGTGGCCCTGTGTGCAATGCGGGAGAGAAAGAGTGGCTCAT 1776
QY 781 GGAATCGTGTGATCCCGGTGGAATGTTCAAGAGTGGCCGTGTGTGAGAGACAGAGTGGCTT 840
DB 1777 GGAATCGTGTGATCCCGGTGGAATGTTCAAGAGTGGCCGTGTGTGAGAGACAGAGTGGCTT 1836
QY 841 CGGTGTGTGATCGGAGAGAGAGAGTGGCTTGTGTGAGAGAGAGAGTGGCTTGG 900
DB 1837 CGGTGTGTGATCGGAGAGAGAGAGTGGCTTGTGTGAGAGAGAGAGTGGCTTGG 1896
QY 901 GGGGCCCCCTTGACCTTTGATGAACAGATGGAGAGAGAGAGAGAGTGGAGAGAC 960

```

Db 1897 GGGGCCCCCTTGACCTTTGATGAACAGATGGAGAGAGAGAGAGAGTGGAGAGAC 1956

QY 961 TGGAAAAATGATTAAGGCGAGACCGCTCAGAGAGCGCAGAGATGGCCCTTAC 1012

Db 1957 TGGAAAAATGATTAAGGCGAGACCGCTCAGAGAGCGCAGAGATGGCCCTTAC 2008

## RESULT 6

CR602561 2070 bp mRNA linear HTC 21-JUL-2004

LOCUS full-length cDNA clone CS01068YN23 of Placenta Cot 25-normalized

DEFINITION of Homo sapiens (human).

ACCESSION CR602561.1 GI:50483368

VERSION HTC; CNSLT\_CDNA.

KEYWORDS Homo sapiens (human)

SOURCE Homo sapiens

ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 2070)

AUTHORS Li, W.B., Gruber, C., Jesse, J., and Polayes, D.

TITLE Full-length cDNA libraries and normalization

JOURNAL Unpublished

REMARK Contact: Peng Liang Email: filiang@lifetech.com URL: http://fulllength.invitrogen.com/Invitrogen Corporation 1600

Faraday Avenue

2 (bases 1 to 2070)

REFERENCE Genoscope.

AUTHORS Direct Submission

TITLE Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage :

JOURNAL BP 191 91006 EVRY cedex - FRANCE (E-mail: seqref@genoscope.cns.fr

- Web : www.genoscope.cns.fr)

COMMENT 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime

end enriched, double-strand cDNA was digested with Not I and cloned

into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library

was normalized. Library was constructed by Life Technologies, a

division of Invitrogen.

FEATURES Location/Qualifiers

source 1. 2070

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon.9606"

/clone="CS01068YN23"

/issue\_type="Placenta Cot 25-normalized"

/plasmid="PCMVSPORT\_6"

## ORIGIN

Query Match 98.4%; Score 996; DB 3; Length 2070;

Best Local Similarity 98.4%; Pred. No. 7.8e-237; Mismatches 12; Indels 0; Gaps 0;

Matches 1000; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

```

QY 1 AGGACCGCGTGAATGGGACCGCTGGAGAGCGAGTGGCTTCAATAGCCTGGTGGACC 60
DB 942 AGGACCGCGTGAATGGGACCGCTGGAGAGCGAGTGGCTTCAATAGCCTGGTGGACC 1001
QY 61 CATGATGAAGACACAGATCTTGAATGAAGCCACTGTAGATCAGATGAAGACTCTGA 120
DB 1002 CATGATGAAGACACAGATCTTGAATGAAGCCACTGTAGATCAGATGAAGACTCTGA 1061
QY 121 CAACAGTGCATTTATATACAGATTTAATGACAGTGTGACTTATATATGATCTGGCAGA 180
DB 1062 CAACAGTGCATTTATATACAGATTTAATGACAGTGTGACTTATATATGATCTGGCAGA 1121
QY 181 CTCTTTAAGCAGTGTGGGTTGTATAGATGAACAAGAACTGGGACCACTGATCCA 240
DB 1122 CTCTTTAAGCAGTGTGGGTTGTATAGATGAACAAGAACTGGGACCACTGATCCA 1181
QY 241 CATCTAAGTGAACAGAGAAACAGAAAGCCCAAGCGATGCCAAGTGTCTATGAAGA 300
DB 1182 CATCTAAGTGAACAGAGAAACAGAAAGCCCAAGCGATGCCAAGTGTCTATGAAGA 1241
QY 301 CCCACCACTGCGCAAGGCTGCGTGGATGATGTTGATGGAAAGATTTTCAAGGAGCAA 360

```

Db 1242 CCCACCACTGCGCAAGGCTGCGGTGAAATGTTGATGGGAAAGATTTTCAAGGAGCA 1301  
 QY 361 ACTTAAGTCTCCCTTGTCTCGGAAGAAGCTTCCATGAAGATGCGGGGTCTGCC 420  
 Db 1302 ACTTAAGTCTCCCTTGTCTCGGAAGAAGCTTCCATGAAGATGCGGGGTCTGCC 1361  
 QY 421 ACCCGTGAAGGAGAGGAGCATGCAACCACTCCGTGAGTCCAGAGGCCAGAGG 480  
 Db 1362 ACCCGTGAAGGAGAGGAGCATGCAACCACTCCGTGAGTCCAGAGGCCAGAGG 1421  
 QY 481 TCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 540  
 Db 1422 TCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1481  
 QY 541 AAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 600  
 Db 1482 AAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1541  
 QY 601 TGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 660  
 Db 1542 TGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1601  
 QY 661 GTGCAACAGTGTAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 720  
 Db 1602 GTGCAACAGTGTAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1661  
 QY 721 GGGTGTGATCGTGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 780  
 Db 1662 GGGTGTGATCGTGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1721  
 QY 781 GATGCTGTGTGTCTCCGAGTGAATGTTCAAGAGGAGGAGGAGGAGGAGGAGGAGG 840  
 Db 1722 GATGCTGTGTGTCTCCGAGTGAATGTTCAAGAGGAGGAGGAGGAGGAGGAGGAGG 1781  
 QY 841 CCGTGTGTGTGTCTCCGAGTGAATGTTCAAGAGGAGGAGGAGGAGGAGGAGGAGG 900  
 Db 1782 CCGTGTGTGTGTCTCCGAGTGAATGTTCAAGAGGAGGAGGAGGAGGAGGAGGAGG 1841  
 QY 901 GGGGCCCCCTTGACCTTTGATGGAACAGATGGAGAGGAGGAGGAGGAGGAGGAGG 960  
 Db 1842 GGGGCCCCCTTGACCTTTGATGGAACAGATGGAGAGGAGGAGGAGGAGGAGGAGG 1901  
 QY 961 TGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1012  
 Db 1902 TGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1953  
 RESULT 7  
 CR594868  
 LOCUS full-length cDNA clone CS0D1064YE15 of Placenta Cot 25-normalized 2093 bp mRNA HTc 21-JUL-2004  
 DEFINITION of Homo sapiens (human).  
 ACCESSION CR594868  
 VERSION CR594868.1 GI:50475675  
 KEYWORDS HTc; CNSLT cDNA.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 2093)  
 AUTHORS Li, W. B., Gruber, C., Jesse, J., and Polayes, D.  
 TITLE Full-length cDNA libraries and normalization  
 JOURNAL Unpublished  
 REMARK Contact: Feng Liang Email: fliang@lifetech.com URL: http://fulllength.invitrogen.com/Invitrogen Corporation 1600 Faraday Avenue  
 REFERENCE 2 (bases 1 to 2093)  
 AUTHORS Genoscope.  
 TITLE Direct Submission  
 JOURNAL Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail: seque@genoscope.cns.fr)  
 COMMENT - Web : www.genoscope.cns.fr  
 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime

end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies, a division of Invitrogen.  
 Location/Qualifiers  
 source  
 1.2093  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="CS0D1064YE15"  
 /issue\_type="Placenta Cot 25-normalized"  
 /plasmid="pCMVSPORT\_6"  
 ORIGIN  
 Query Match 98.4%; Score 996; DB 3; Length 2093;  
 Best Local Similarity 98.8%; Pred. No. 7.8e-237;  
 Matches 1000; Conservative 0; Mismatches 12; Indels 0; Gaps 0;  
 1 AAGAGCGGAGTGAATGGAGCGCTGAGAGCGAGGTGGCTTCAATAGCTGTGAGC 60  
 Db 942 AAGAGAGCGGAGTGAATGGAGCGCTGAGAGCGAGGTGGCTTCAATAGCTGTGAGC 1001  
 QY 61 CATGATGAAGACCAAGATCTTGAATAGAGCCCACTGTAGATCCAGATGAAGACTCTGA 120  
 Db 1002 CATGATGAAGACCAAGATCTTGAATAGAGCCCACTGTAGATCCAGATGAAGACTCTGA 1061  
 QY 121 CAACAGTGAATTATGTAAGAGTGAATGAAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 180  
 Db 1062 CAACAGTGAATTATGTAAGAGTGAATGAAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 1121  
 QY 181 CTTCCTTAAGCAGTGTGGGTTGTTAAGATGAACAGAGAACTGGCAACCCATGATCCA 240  
 Db 1122 CTTCCTTAAGCAGTGTGGGTTGTTAAGATGAACAGAGAACTGGCAACCCATGATCCA 1181  
 QY 241 CATCTACTGACAGAGAAACAGAGAAAGCCCAAGGCGATGCCACAGTGTCTTATGAAGA 300  
 Db 1182 CATCTACTGACAGAGAAACAGAGAAAGCCCAAGGCGATGCCACAGTGTCTTATGAAGA 1241  
 QY 301 CCCACCACTGCGCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 360  
 Db 1242 CCCACCACTGCGCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1301  
 QY 361 ACTTAAGTCTCCCTTGTCTCGGAAGAAGCTTCCATGAAGATGCGGGGTCTGCC 420  
 Db 1302 ACTTAAGTCTCCCTTGTCTCGGAAGAAGCTTCCATGAAGATGCGGGGTCTGCC 1361  
 QY 421 ACCCGTGAAGGAGAGGAGCATGCAACCACTCCGTGAGTCCAGAGGCCAGAGG 480  
 Db 1362 ACCCGTGAAGGAGAGGAGCATGCAACCACTCCGTGAGTCCAGAGGCCAGAGG 1421  
 QY 481 TCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 540  
 Db 1422 TCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1481  
 QY 541 AAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 600  
 Db 1482 AAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1541  
 QY 601 TGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 660  
 Db 1542 TGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1601  
 QY 661 GTGCAACAGTGTAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 720  
 Db 1602 GTGCAACAGTGTAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1661  
 QY 721 GGGTGTGATCGTGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 780  
 Db 1662 GGGTGTGATCGTGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1721  
 QY 781 GATGCTGTGTGTCTCCGAGTGAATGTTCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGG 840  
 Db 1722 GATGCTGTGTGTCTCCGAGTGAATGTTCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1781



OY		841	CCGTGATGGCGGGGCATAGACCAAGAGTGGCTTTGTGGAGGAACAAGAGTGCCCTMG	900
Db		1762	CCGTGATGGCGGGGCATAGACCAAGAGTGGCTTTGTGGAGGAACAAGAGTGCCCTMG	1841
OY		901	GCGGCCCCCTGGAACCTTTGATGAAACAGATGGAGAAGAAAGAGAGACGTGAGAGACC	960
Db		1842	GCGGCCCCCTGGAACCTTTGATGAAACAGATGGAGAAGAAAGAGAGACGTGAGAGACC	1901
OY		961	TGGAAAAATGATTAAAGCGAGACACCGTCAGAGCGCCAGAGATCGGCCCTTAC	1012
Db		1902	TGGAAAAATGATTAAAGCGAGACACCGTCAGAGCGCCAGAGATCGGCCCTTAC	1953
RESULT 8				
LOCUS	CR610888			
DEFINITION	Full-length cDNA clone CSDFP037YI07 of Fetal brain of Homo sapiens (human).	2126 bp	mRNA	linear HTC 21-JUL-2004
ACCESSION	CR610888			
VERSION	CR610888.1			GI:50491695
KEYWORDS	HTC; CNSLT_cDNA.			
SOURCE	Homo sapiens (human)			
ORGANISM	Homo sapiens			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
AUTHORS	Li, W.B., Gruber, C., Jeesee, J. and Polayes, D.			
JOURNAL	Unpublished			
REMARK	Contact : Feng liang Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com/ Invitrogen Corporation 1600 Faraday Avenue 2 (bases 1 to 2126)			
REFERENCE	Genoscope.			
AUTHORS	Direct Submission			
JOURNAL	Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : sequefegenoscope.cns.fr - Web : www.genoscope.cns.fr)			
COMMENT	1st strand cDNA was primed with a NotI-oligo(dT) primer. Five primers end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by life technologies, a division of invitrogen.			
FEATURES	Location/Qualifiers			
SOURCE	1..2126			
	/organism="Homo sapiens"			
	/mol_type="mRNA"			
	/db_xref="taxon:9606"			
	/clone="CSDFP037YI07"			
	/tissue.type="Fetal brain"			
	/plasmid="pCMVSPORT_6"			
ORIGIN				
Query Match	98.4%; Score 996; DB 3; Length 2126;			
Best Local Similarity	98.8%; Pred. No. 7.8e-237;			
Matches 1000; Conservative	0; Mismatches 12; Indels 0; Gaps 0;			
OY		1	AGGAGCGCGGTGAGATGGGACGCGCTGGAGAGACGAGTGGCTTCAATAAGCCTGTGTGACC	60
Db		942	AGGAGGACGCGGTGAGATGGGCGCTGGAGAGACGAGTGGCTTCAATAAGCCTGTGTGACC	1001
OY		61	CATGATGAAGAACACAGATCTTGATCTTAGCGCCACCTGTAGATCCAGATGAAGACTCTGA	120
Db		1002	CATGATGAAGAACACAGATCTTGATCTTAGCGCCACCTGTAGATCCAGATGAAGACTCTGA	1061
OY		121	CAAAGTGCATTTATGTACAAGATTAAATGACAGTGTGACTCTAAGTAGTCTGGCAGA	180
Db		1062	CAAAGTGCATTTATGTACAAGATTAAATGACAGTGTGACTCTAAGTAGTCTGGCAGA	1121
OY		181	CTTCTTTAAGAGAGTGGGGTTGTTAATATGAACAAGAACTGGGCAACCCATGATCCA	240
Db		1122	CTTCTTTAAGAGAGTGGGGTTGTTAATATGAACAAGAACTGGGCAACCCATGATCCA	1181

QY	241	UATCTACTGGACAAGAGAAACAAGAAAGCCCAAGAGCGATATCCACAGTGTCTCTATGAAGA	300
Db	1182	CATCTACTGGACAAGAGAAACAAGAAAGCCCAAGAGCGATATCCACAGTGTCTCTATGAAGA	12411
QY	301	CCCAACCACTGCTCCAAAGCGTGCCTGTGAATGAGTTTGTATGGGAAAGATTTTCAAGGAGCAA	360
Db	1242	CCCAACCACTGCTCCAAAGCGTGCCTGTGAATGAGTTTGTATGGGAAAGATTTTCAAGGAGCAA	13010
QY	361	ACTTAAAGTCTCCCTTGTCTGGAGAAAGCTTCAATGAACATATGCGGGGTGCTTGGC	420
Db	1302	ACTTAAAGTCTCCCTTGTCTGGAGAAAGCTTCAATGAACATATGCGGGGTGCTTGGC	13610
QY	421	AACCCGTGAAGGCGAGAGCATATGCCAACCATCTCCGTGAAGTATCCAGAGGCCCAAGAGG	480
Db	1362	AACCCGTGAAGGCGAGAGCATATGCCAACCATCTCCGTGAAGTATCCAGAGGCCCAAGAGG	14210
QY	481	TCCTGGAGGGAACCAATGGGTCGATGGAGAGCGCGTGAAGAGATAGAGAGGCTTCCCTAC	540
Db	1422	TCCTGGAGGGAACCAATGGGTCGATGGAGAGCGCGTGAAGAGATAGAGAGGCTTCCCTAC	14810
QY	541	AAGAGAGACCCCGGGGTTCCCGAGGGAACCCCTCTGAGAGAGAAACGTCCACACCGAGC	600
Db	1482	AAGAGAGACCCCGGGGTTCCCGAGGGAACCCCTCTGAGAGAGAAACGTCCACACCGAGC	15410
QY	601	TGAGAGCTGGAGAGTCCCAATCCGGGTGTGGAACCAAGAACTTCCGTGAAGAACAGA	660
Db	1542	TGAGAGCTGGAGAGTCCCAATCCGGGTGTGGAACCAAGAACTTCCGTGAAGAACAGA	16010
QY	661	GTGCACACAGATGTAAAGGCCCAAGACCTGAAAGGCTTCTCCCGCACCTTTCCGCCCC	720
Db	1602	GTGCACACAGATGTAAAGGCCCAAGACCTGAAAGGCTTCTCTCCCGCACCTTTCCGCCCC	16610
QY	721	GCGTGTGATGTGTGGCAGAGGTGGCCCTGGTGCATGCGGGAGGAAAGAGGTGGCTCAT	780
Db	1662	GCGTGTGATGTGTGGCAGAGGTGGCCCTGGTGCATGCGGGAGGAAAGAGGTGGCTCAT	17210
QY	781	GGATCGTGTGTGTCCTCCGGTGGAAATGTTCAAGAGTGGCCGTGTGGAGACAGAGGTGGCTT	840
Db	1722	GGATCGTGTGTGTCCTCCGGTGGAAATGTTCAAGAGTGGCCGTGTGTGGAGACAGAGGTGGCTT	17810
QY	841	CCGTGTGTGGCCGGGGCATATGACCCGAGGTGGCTTGTGTGAGAGAAAGAGGTGGCCCTGG	900
Db	1782	CCGTGTGTGGCCGGGGCATATGACCCGAGGTGGCTTGTGTGTGGAGAAAGAGGTGGCCCTGG	18410
QY	901	GGGGCCCCCTTGACCTTTGATGAAACAATGTGGAGAAAGAAAGAGAGAGAGCTGGAGAC	960
Db	1842	GGGGCCCCCTTGACCTTTGATGAAACAATGTGGAGAAAGAAAGAGAGAGAGCTGGAGAGAC	19010
QY	961	TGGAAAAATGGATTAAGAGCGAGCACCGTCAAGGAGCGAGAGATGGCGCTTAC	1012
Db	1902	TGGAAAAATGGATTAAGAGCGAGCACCGTCAAGGAGCGAGAGATGGCGCTTAC	1953
RESULT 9			
CR625247			
LOCUS full-length cDNA clone CS0DB0051F11 of Placenta of Homo sapiens			
DEFINITION (human).			
ACCESSION CR625247			
VERSION CR625247.1 GI:50506054			
KEYWORDS HTC; CNSLT cDNA.			
SOURCE Homo sapiens (human)			
ORGANISM Homo sapiens			
REFERENCE Eukaryotic; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
AUTHORS Mammalia; Euteleostomi; Primates; Catarrhini; Homnidae; Homo.			
TITLE Li, W.B., Gruber, C., Jesse, J., and Polayes, D.			
JOURNAL Full-length cDNA libraries and normalization			
REMARK Unpublished			
Contact: Feng Liang Email: fliang@life.techn.com URL: http://fulllength.invitrogen.com/Invitrogen Corporation 1600			
Fairway Avenue			



QY 121 CAAAGTGCATTTATGTACAGAGATTAATAAGACAGTGTGACTCTAGATGATCGGAGA 180  
 DB 1092 CAACAGTGCATTTATGTACAGAGATTAATAAGACAGTGTGACTCTAGATGATCGGAGA 1151  
 QY 181 CTTCTTTAGCAGTGTGGGGTGTGTAAGATGAAGAACAGAGAACTGGGAGACCCATGATCA 240  
 DB 1152 CTTCTTTAGCAGTGTGGGGTGTGTAAGATGAAGAACAGAGAACTGGGAGACCCATGATCA 1211  
 QY 241 CATCTACTGACAGAGAAACAGAGAAAGCCAAAGGCGATGCCAGTGTCTTATGAAGA 300  
 DB 1212 CATCTACTGACAGAGAAACAGAGAAAGCCAAAGGCGATGCCAGTGTCTTATGAAGA 1271  
 QY 301 CCCACCACTGCCAGGCTGCGCGATGATGATGGGAAAGATTTTCAAGGAGACCA 360  
 DB 1272 CCCACCACTGCCAGGCTGCGCGATGATGATGGGAAAGATTTTCAAGGAGACCA 1331  
 QY 361 ACTTAAAGTCTCCCTTCTCGAGAGAAAGCTTCAATGAACAGTATGCGGGGTGTCTGCC 420  
 DB 1332 ACTTAAAGTCTCCCTTCTCGAGAGAAAGCTTCAATGAACAGTATGCGGGGTGTCTGCC 1391  
 QY 421 ACCCGGTGAGGCGAGAGGATGCCACCACTCCGTGAGGTCCAGAGGCCAGAGAG 480  
 DB 1392 ACCCGGTGAGGCGAGAGGATGCCACCACTCCGTGAGGTCCAGAGGCCAGAGAG 1451  
 QY 481 TCCGTGGGGGACCCATGGGTGCGCATGGAGGCGGTGAGAGAGATAGAGAGGCTTCCCTCC 540  
 DB 1452 TCCGTGGGGGACCCATGGGTGCGCATGGAGGCGGTGAGAGAGATAGAGAGGCTTCCCTCC 1511  
 QY 541 AAGAGACCCCGGGGGTTCGCGAGAGAAACCCCTCTGAGAGAGAAACCTCCAGACCGAGC 600  
 DB 1512 AAGAGACCCCGGGGGTTCGCGAGAGAAACCCCTCTGAGAGAGAAACCTCCAGACCGAGC 1571  
 QY 601 TGAAGACTGCGAGTGTCCCAATCCGGGTGTGGAACCAAGAACTTGCTGAGAGACAGA 660  
 DB 1572 TGAAGACTGCGAGTGTCCCAATCCGGGTGTGGAACCAAGAACTTGCTGAGAGACAGA 1631  
 QY 661 GTGCAACCACTGTAAAGCCCAAAAGCTGAAAGGCTTCTCCGCAACCTTCCGCCCC 720  
 DB 1632 GTGCAACCACTGTAAAGCCCAAAAGCTGAAAGGCTTCTCCGCAACCTTCCGCCCC 1691  
 QY 721 GGGTGTGATCTGTGCGAGAGTGGCCCTGTGTCGATCGGGGAGAGAGAGTGGCTCAT 780  
 DB 1692 GGGTGTGATCTGTGCGAGAGTGGCCCTGTGTCGATCGGGGAGAGAGAGTGGCTCAT 1751  
 QY 781 GGATCGTGTGTTCGCGGTGGAATGTTCAAGAGTGGCCGTGTGAGAGACAGAGTGGCTT 840  
 DB 1752 GGATCGTGTGTTCGCGGTGGAATGTTCAAGAGTGGCCGTGTGAGAGACAGAGTGGCTT 1811  
 QY 841 CCGTGTGTCGCGGGGCGATGAGACCGAGTGGCTTGTGTGAGAGAGAGAGTGGCTTGG 900  
 DB 1812 CCGTGTGTCGCGGGGCGATGAGACCGAGTGGCTTGTGTGAGAGAGAGAGTGGCTTGG 1871  
 QY 901 GGGGCCCC 908  
 DB 1872 GGGGCCCC 1879

RESULT 11  
 AK014366 1379 bp mRNA 1linear HTC 03-Apr-2004  
 LOCUS AK014366  
 DEFINITION Mus musculus 17 days embryo head cDNA, RIKEN full-length enriched library, clone:3300002D11 product:EWing sarcoma homolog, full insert sequence.  
 ACCESSION AK014366  
 VERSION AK014366.1 GI:12852167  
 KEYWORDS HTC; CAP trapper.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 REFERENCE 1 Carninci, P. and Hayashizaki, Y.

TITLE High-efficiency full-length cDNA cloning  
 JOURNAL Meth. Enzymol. 303, 19-44 (1999)  
 MEDLINE 99279253  
 PUBMED 10349636  
 REFERENCE 2  
 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Komno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y. Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes  
 JOURNAL Genome Res. 10 (10), 1617-1630 (2000)  
 MEDLINE 20499374  
 PUBMED 11042159  
 REFERENCE 3  
 Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Komno, H., Akiyama, J., Nishi, K., Kishimoto, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiiwake, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watabiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsunura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y. RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer  
 JOURNAL Genome Res. 10 (11), 1757-1771 (2000)  
 MEDLINE 20530913  
 PUBMED 11076861  
 REFERENCE 4  
 The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.  
 TITLE Functional annotation of a full-length mouse cDNA collection  
 JOURNAL Nature 409, 685-690 (2001)  
 MEDLINE 117571771  
 PUBMED 11076861  
 REFERENCE 5  
 The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.  
 TITLE Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs  
 JOURNAL Nature 420, 563-573 (2002)  
 MEDLINE 12002120  
 PUBMED 12002120  
 REFERENCE 6 (bases 1 to 1379)  
 Adachi, T., Aizawa, K., Akahira, S., Akimura, T., Arai, A., Aono, H., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Fukunishi, Y., Furuno, M., Hanagaki, T., Hara, A., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hori, F., Imocani, K., Ishii, Y., Itoh, M., Izawa, M., Kasukawa, T., Kato, H., Kawai, J., Kojima, Y., Komno, H., Kouda, M., Koya, S., Kuribara, C., Matsuyama, T., Miyazaki, A., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Okazaki, Y., Okido, T., Owa, C., Saito, H., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki, D., Shibata, K., Shibata, Y., Shinagawa, A., Shiraki, T., Segabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F., Tanaka, T., Tejima, Y., Toya, T., Yamamura, T., Yasunishi, A., Yoshida, K., Yoshino, M., Muramatsu, M. and Hayashizaki, Y.  
 TITLE Direct Submission  
 JOURNAL Submitted (10-Jul-2000) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.jp, URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)  
 COMMENT Please visit our web site (http://genome.gsc.riken.jp/) for further details.  
 CDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in Riken. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. First strand cDNA was primed with a primer [5'-GAGAGAGAGATCCAGATCCAGAGCTCTTTTCTTTTCTTTT-3']. cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5'-GAGAGAGATTCCTGAGTAAATTAATATCCCTCCCTCCCTCC-3']. cDNA was cleaved with XhoI and SstI. Cloning sites, 5' end: XhoI, 3' end: SstI. Host: SOLR.  
 FEATURES  
 source Location/Qualifiers  
 1. 1379

```

/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM_DB:3300002D11"
/db_xref="taxon:10090"
/clone="3300002D11"
/issue_type="head"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="17 days embryo"
<1..1004
/notes="unlabeled protein product; Ewing sarcoma homolog
(MGD|MG1:99960, GB|NM_007968, evidence: BLASTN, 99%,
match=2172)
putative"
/codon_start=3
/protein_id="BAB29301.1"
/db_xref="GI:12852168"
/translation="LQSGAGRGGRGKPKGPMDEGPDLDLPIIDPEDSDNSAIYVG
LNDNVTLDLADLPKOCGVYKMKRGTGQPMIHTYLDKRGKPKGDAIVSYEDPTAKA
AVDFPGKDFQSKLYSLAKKPPKMSNMGMPPRBGRGMPPLIAGRGPGPGPGP
MGRMGGRGDRGSPPRGPRGSRNBSGSGNVDRAGDWCPNPGGNOFARTCN
QCKAPKPEGLPPPPPPGDRGRGGRGGLMDRGGMGRGGRGGRGGRGGRGGRG
RGRGMDRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGR
RGRGMDRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGR"
1351..1356
/notes="putative"
1379
/notes="putative"

```

polysignal

polysite

ORIGIN

Query Match 87.2%; Score 882.6; DB 3; Length 1379;  
 Best Local Similarity 92.6%; Pred. No. 1.2e-208;  
 Matches 927; Conservative 0; Mismatches 74; Indels 0; Gaps 0;

```

12 GAATGGGACGCGTGGAGAGGAGTGGCTTCAATTAAGCTGTGACCCATGATGAG 71
1 GACTGGGACGCGTGGAGAGGAGTGGCTTCAATTAAGCTGTGACCCATGATGAG 60
72 GACCAATCTTGAATCTAGGCCCACTGTATGATCCAGATGAAGACTCGAACAAGTGA 131
61 GACCAATCTTGAATCTAGGCCCTTCTTATGATCCAGATGAAGACTCGAACAAGTGA 120
132 TTTATGTACAAGATTAATGACAGTGTGATCTAGATGATCTGGCAGACTTCTTAA 191
121 TTTATGTACAAGATTAATGACAGTGTGATCTAGATGATCTGGCAGACTTCTTAA 180
192 AGTGTGGGTTTGAATGATGAACAAGAACTGGGCAACCCATGATCAATCTTAC 251
181 AGTGTGGGTTTGAATGATGAACAAGAACTGGGCAACCCATGATCAATCTTAC 240
252 ACAAGGAAACAGAAACCCCAAGGCGATGCCAAGTGTCTTATGAAGACCCCACT 311
241 ATAAAGGAAACAGAAACCCCAAGGCGATGCCAAGTGTCTTATGAAGACCCCACT 300
312 CCAAGGCTGCGGTGAATGATTTTGAAGGAAAGATTTTCAAGGAGCAACTTAA 371
301 CCAAGGCTGCGGTGAATGATTTTGAAGGAAAGATTTTCAAGGAGCAACTTAA 360
372 CCGTGTGCGAAGAGCCTTCAATGAACAGTATGCGGGGTGTCTGCAACCCCGTAG 431
361 CCGTGTGCGAAGAGCCTTCAATGAACAGTATGCGGGGTGTCTGCAACCCCGTAG 420
432 GCAGAGGATGCCCAACCACTCCGTGAGAGTCCAGAGGCGCCAGAGGATCTTGG 491
421 GCAGAGGATGCCCAACCACTCCGTGAGAGTCCGTGAGGCGCCAGAGGATCTTGG 480
492 CCATGGGTGCGATGAGAGCGGTGAGAGGATTAAGAGAGGCTTCCCTCAAGAGAG 551
481 CCATGGGTGCGATGAGAGCGGTGAGAGGATTAAGAGAGGCTTCCCTCAAGAGAG 540
552 GGGGTTCCGAGAGGAGCCCTCTGAGAGAGAGAGAGTTCAGACAGAGCTGAGAG 611
541 GAGGCTCGAGAGAGAGCCCTCTGAGAGAGAGAGAGTTCAGACAGAGCTGAGAG 600

```

```

QY 612 AGTGTCCCAATCGGGTGTGAGAAACAGAACTTGCGCTGAGAAACAGAGTGCACAGT 671
DB 601 AGTGTCCCAATCGGGTGTGAGAAACAGAACTTGCGCTGAGAAACAGAGTGCACAGT 660
QY 672 GTAAAGGCCCCAAAGCCTGAAAGCTTCTCCGCGCACCTTTTCCGCCCGGGGTGTGATC 731
DB 661 GTAAAGGCCCCAAAGCCTGAAAGCTTCTCCGCGCACCTTTTCCGCCCGGGGTGTGATC 720
QY 732 GTGCGAGAGTGGCCCTGTGTCATGCGGGGAGAGAGAGTGGCCCTCAATGATCGTGTG 791
DB 721 GTGCGAGAGTGGCCCTGTGTCATGCGGGGAGAGAGAGTGGCCCTCAATGATCGTGTG 780
QY 792 GTCCCGGTGAATGTTCAAGAGTGGCCGTGTGAGAGAGAGAGTGGCCCTTCCGTGTGCGC 851
DB 781 GTCTGTGAGAGATGTTTCAAGAGTGGCCGTGTGAGAGAGAGAGTGGCCCTTCCGTGTGCGC 840
QY 852 GGGGCAATGACCGAGGTGGCTTTGTGAGAGAGAGAGAGTGGCCCTTGGGGGGCCCTT 911
DB 841 GTGGAATGACCGAGGTGGCTTTGTGAGAGAGAGAGAGTGGCCCTTGGGGGGCCCTT 900
QY 912 GACCTTTGATGAAACAGATGGAGAGAGAGAGAGAGAGTGGAGACCTGGAAGAAATG 971
DB 901 GACCTTTGATGAAACAGATGGAGAGAGAGAGAGAGAGTGGAGACCTGGAAGAAATG 960
QY 972 ATAAAGCGAGACCGCTCAGAGCGCAGAGATGCGCCCTAC 1012
DB 961 ATAAAGCGAGACCGCTCAGAGCGCAGAGATGCGCCCTAC 1001

```

RESULT 12

AK019460

LOCUS

DEFINITION

AK019460 2107 bp mRNA linear HTC 03-Apr-2004

Mus musculus 18 days pregnant adult female placenta and extra embryonic tissue cDNA, RIKEN full-length enriched library, clone:3830417B11 product:Ewing sarcoma homolog, full insert sequence.

AK019460

VERSION

AK019460.1 GI:12859676

KEYWORDS

HTC; CAP trapper

SOURCE

Mus musculus (house mouse)

ORGANISM

Mus musculus

Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE

1 Carninci, P. and Hayashizaki, Y.  
 High-efficiency full-length cDNA cloning  
 Meth. Enzymol. 303, 19-44 (1999)

JOURNAL

99279253

MEDLINE

PUBMED

10349636

REFERENCE

2 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Komno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
 Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes  
 Genome Res. 10 (10), 1617-1630 (2000)

JOURNAL

20499374

MEDLINE

PUBMED

11042159

REFERENCE

3 Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Komno, H., Akiyama, J., Nishi, K., Katsunari, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Mitsuhashi, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Matshiki, M., Itoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, D., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.  
 RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer  
 Genome Res. 10 (11), 1757-1771 (2000)

JOURNAL

20530913

MEDLINE

PUBMED

11076861

REFERENCE

4 The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.

**TITLE** Functional annotation of a full-length mouse cDNA collection  
**JOURNAL REFERENCE** Nature 409, 685-690 (2001)  
**AUTHORS** 5 The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.  
**TITLE** Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs  
**JOURNAL REFERENCE** Nature 420, 563-573 (2002)  
**AUTHORS** 6 (bases 1 to 2107) Adachi, J., Aizawa, K., Akahira, S., Akimura, T., Arai, A., Aono, H., Arkawa, T., Bono, H., Carninci, P., Fukuda, S., Furukoshi, Y., Furuno, M., Hanagaki, T., Hara, A., Hayatsu, N., Hiramoto, K., Hirooka, T., Hori, F., Imotani, K., Ishii, Y., Itch, M., Izawa, M., Kasukawa, T., Kato, H., Kawai, Y., Koizumi, Y., Komori, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Niimi, K., Nomura, K., Numasaki, R., Ohno, M., Okazaki, Y., Okido, T., Owa, C., Salto, H., Salto, R., Sakai, C., Sakai, K., Sano, H., Saeki, D., Shibata, K., Shibata, Y., Shingawa, A., Shiraki, T., Sogabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takehashi, F., Tanaka, T., Teijima, Y., Toyai, T., Yamamura, T., Yasunishi, A., Yoshida, K., Yoshino, M., Muramatsu, M. and Hayashizaki, Y.

**TITLE** Direct Submission  
**JOURNAL** Submitted (18-AUG-2000) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsunumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gs.c.riken.jp, URL: http://genome-gsc.riken.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)

**COMMENT** Please visit our web site (<http://genome.gsc.riken.jp/>) for further details.  
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. First strand cDNA was primed with a primer [5'-GAAGAGAAGATCCAAAGAGCTTTTATTATTTTAA 3'], cDNA was prepared by using triethanolamine thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 10.0 and subtraction to Rot = 50.0. Second strand cDNA was prepared with the primer adapter of sequence [5'-GAAGAGAAGATTCGAGTTAATTAAATTATCCCCCCCCC 3']. cDNA was cleaved with BamHI and XhoI. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLX I. Cloning sites, 5' end: SalI, 3' end: BamHI. Host: DH10B.

**FEATURES**  
**SOURCE** Location/Qualifiers  
1..2107  
    organism="Mus musculus"  
    mol\_type="mRNA"  
    strain="C57BL/6J"  
    db\_xref="PANTOM.DB:1830417B11"  
    db\_xref="taxon:10090"  
    clone="3830417B11"  
    sex="female"  
    tissue\_type="Placenta and extra embryonic tissue"  
    clone\_lib="RIKEN full-length enriched mouse cDNA library"  
    dev\_stage="18 days pregnant adult"  
17..1984  
    name="unnamed protein product; Ewing sarcoma homolog (MGI:99960, GB|NW\_007968, evidence: BLASTN, 99%, match=2172)  
    putative"  
    codon\_start=1  
    protein\_id="BAB31732.1"  
    db\_xref="GI:12859677"  
    translation="MASTDIYVTSQAAAQGVSAVTAPTOGYAQTDAVGSGSYRT  
GGPDIVTAQAQTAVGTQAVATVSQPGSPPTPAQASQPVQSGYGTADSTT  
ATVTTQMSVAAGTAVGTPYVGOOPDTATPRADKPKAESDPOSSNGYNQ  
PSLGSQNSVSPDVPPGSPYMPAQITATPASPYSVSSPOPNQDSISQONTYGCS  
SYQQSSSYGQSSSYGQPPTYSIPQTSYSISAPQYSIQSSSYSQSSSFQNDHPSNG  
VYGEISGFSGPGENRSLISGDNRGRGDRGMRGGRGGGGLGAGEKGKFK  
PGMGSDPDLIDLPIDPEDSDNSAIYYQGLANDVTLLDFPFQCGVVNNKKTK"

Query Match	86.9%	Score 879.2	DB 3	Length 2107
Best Local Similarity	91.7%	Pred. No. 9.4e-208		
Matches 928	Conservative	0	Mismatches 84	Indels 0
Gaps	0			
Qy 1	AGGACGCGGTGGAATGGGCGACGCGCTGGAGAGCGAGTGGCTTCAATAAAGCTGGTGGACC	60		
Db 970	AGGAGAGACGCGGTGGAATGGGCGCGCTGGAGAGCGAGTGGCTTCAATAAAGCTGGTGGACC	1029		
Qy 61	CATGAGATGAAGACCAAGATCTTGAATCTTAAGGCCACTGTAGATCCAGATGAAGAAGCTCTGA	120		
Db 1030	CATGAGATGAAGACCAAGATCTTGAATCTTAAGGCCACTGTAGATCCAGATGAAGAAGCTCTGA	1089		
Qy 121	CAACAGTGCATTTATGTACAAGGATTTAAATGACAGTGTGACTCTAGATGATCTGGCAGA	180		
Db 1090	CAACAGTGCATTTATGTACAAGGATTTAAATGACAGTGTGACTCTAGATGATCTGGCAGA	1149		
Qy 181	CTTCTTTAAGCAGTGTGGGGTTGTTAAGATGAACAAGAACTGGGCAACCCATGATCCA	240		
Db 1150	CTTCTTTAAGCAGTGTGGGGTTGTTAAGATGAACAAGAACTGGGCAACCCATGATCCA	1209		
Qy 241	CATCTACCTGTGACAAGGAAACAGGAAAGCCAAAGGGGATGCGCACAGTGTCTTATGAGA	300		
Db 1210	TATCTACTGTGATTAAGAGACAGGAAAGCTTAAGGGGAGCCACAGTGTCTTATGAGA	1289		
Qy 301	CCCAACCACTGCCAAGGCTGCGCTGTGAATGTTGATGAGGAAAGATTTTCAAGGAGCA	360		
Db 1270	TCCAACCACTGCCAAGGCTGCGCTGTGAATGTTGATGAGGAAAGATTTTCAAGGAGCA	1329		
Qy 361	ACTTAAAGTCTCCCTTGTGCGAAAGAAAGCTTCCATGAACAATATGCGGGGTGTCTGCC	420		
Db 1330	ACTTAAAGTGTCTCTTGTCCCGAAAGAAAGCTTCCATGAACAAGTGTGGGGAGGCAATGCC	1389		
Qy 421	ACCCCGTAGGGGCGAGAGGACATGCCACACATCCCGTAGAGGTCCAGAGAGGCCACGAGAG	480		
Db 1390	ACCTGTGAGGAGCAGGGGAGATGCCACACATCTTCGTGAGAGTCTCTGTGGCCAGAGAG	1449		
Qy 481	TCTGTGGGGGACCCATGGGATGCGCATGGGAGGCGGTGAGAGATAGAGAGGCTTCCCTCC	540		
Db 1450	CCCTGAGAGAACCCATGGGATGCGCATGGGAGGCGGTGAGAGAGACAGAGGGGCTTCCCTCC	1509		
Qy 541	AAGAGAACCCCGGGGCTTCCGAGGGAAACCCCTCTGAGAGAGGAAAGCTCCAGACCGAGC	600		
Db 1510	AAGAGGGGCCCCGAGAGGCTTCCAGAGGAAACCCCTCTGAGAGAGGAAATGTCCAGACCGAGC	1569		
Qy 601	TGGAGACTGGGAGGTGTCCCAATCCGGGCTGTGGAAACCAAACTTGTCTTGGAAACAGA	660		
Db 1570	TGGAGACTGGGAGGTGTCCCAATCCGGGCTGTGGAAACCAAACTTGTCTTGGAAACAGA	1629		
Qy 661	GTGACACCAAGTATAGGAGCCCAAGAGCTGGAAGCTTCTCCCGGCAACCTTTCCGAGCCC	720		
Db 1630	ATGCAACCAAGTATAGGAGCCCTTAAGCCCGAGGCTTCTCTCCGCAACCTTTCCACCTCC	1689		
Qy 721	GGGTGTGATCTGTGGCAGAGGTGGCCCTGTGTGCATCCGAGGAGGAAAGAGGTGGCTCAT	780		
Db 1690	GGGTGTGATCTGTGGCAGAGGTGGCCCTGTGTGCATCCGAGGAGGAAAGAGGTGGCTCAT	1749		
Qy 781	GGATCTGTGTGTGTCTCCCGTGGAAATGTTCACAGGTGTGCGGTGTGTGAGACAGAGGTGGCTT	840		
Db 1750	GGACCTGTGTGTGTCTTGTGAGGAAATGTTCACAGGTGTGCGGTGTGTGAGACAGAGGTGGCTT	1809		
Qy 841	CCGTGTGTGTGCGGAGGAGCAATGAGGCTTGTGTGTGTGAGAGAGCAAGATGTGGCCCTTGG	900		
Db 1810	CCGAGGTGTGTGCGGAGGAGCAATGAGGCTTGTGTGTGTGAGAGAGCAAGATGTGGCTTGG	1869		
Qy 901	GGGGCCCCCTTGAGACTTTTGAATGGAACGATGGAGAGGAAAGAGAGGAGCTGTGAGGACC	960		



Db	1870	GGGGGCTCTGGAACCTTTAAATGGAACAGATGGGAGGAAGAGCGGAGCTGGAGGAC	1992
Qy	961	TGGAATAATGATTAAGCGGACACCTTCAGAGCGGAGAGATGGCCCTTAC	1012
Db	1930	TGGGAAATGATTAAGCGGACACCTTCAGAGGACGAGAGCCGGCCCTTAC	1981
RESULT 13			
LOCUS	AK034755	2373 bp	mRNA
DEFINITION	Mus musculus 12 days embryo embryonic body between diaphragm region and neck cDNA, RIKEN full-length enriched library, clone:9430031M11		
ACCESSION	AK034755.1	GI:26084174	
VERSION	AK034755.1	GI:26084174	
KEYWORDS	HTC; CAP trapper.		
SOURCE	Mus musculus (house mouse)		
ORGANISM	Mus musculus		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.		
AUTHORS	Carninci, P. and Hayashizaki, Y.		
TITLE	High-efficiency full-length cDNA cloning		
JOURNAL	High. Enzymol. 303, 19-44 (1999)		
MEDLINE	99279253		
PUBMED	10349636		
REFERENCE			
AUTHORS	Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.		
TITLE	Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes		
JOURNAL	Genome. Res. 10 (10), 1617-1630 (2000)		
MEDLINE	2049374		
PUBMED	11042159		
REFERENCE			
AUTHORS	Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Kitsuana, T., Tashiro, H., Itoh, M., Suni, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kaishiwagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Matshiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.		
TITLE	RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer		
JOURNAL	Genome. Res. 10 (11), 1757-1771 (2000)		
MEDLINE	20530913		
PUBMED	11076861		
REFERENCE			
AUTHORS			
TITLE	The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.		
JOURNAL	Functional annotation of a full-length mouse cDNA collection		
REFERENCE	Nature 409, 685-690 (2001)		
AUTHORS			
TITLE			
JOURNAL			
REFERENCE			
AUTHORS			
TITLE	The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.		
JOURNAL	Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs		
REFERENCE	Nature 420, 563-573 (2002)		
AUTHORS	6 (bases 1 to 2373)		
TITLE	Adachi, J., Aizawa, K., Akiyama, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Haneigaki, T., Haru, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hirooka, T., Hirozane, T., Horii, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Kato, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koyas, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Okazaki, Y., Saito, K., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shingawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toyu, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.		
TITLE	Direct Submission		
JOURNAL	Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of		

	COMMENT	<p>Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration and Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:genome-res@gscc.riken.jp, URL:http://genome.gsc.riken.jp/, Tel:81-45-503-9216, Fax:81-45-503-9216)</p> <p>cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.</p> <p>Please visit our web site for further details.</p> <p>URL:http://genome.gsc.riken.jp/ URL:http://fantom.gsc.riken.jp/ Location/Qualifiers</p>					
	FEATURES						
	source	1..2373 /organism="Mus musculus" /mol_type="mRNA" /strain="C57BL/6J" /db_xref="PANTOM_DB:943003IM11" /db_xref="taxon:10090" /clone="943003IM11" /tissue_type="embryonic body between diaphragm region and neck" /clone_1lb="RIKEN full-length enriched mouse cDNA library" /dev_stage="12 days embryo" 43..2011 /note="Bwing sarcoma homolog [MGI:MGI:99960, GB NM_007968, evidence: BLASTN, 9%, match=2172] putative" 2358..2363 /note="putative" 2373 /note="putative"					
	ORIGIN						
	Query Match	86.6%; Score 876; DB 3; Length 2373;					
	Best Local Similarity	91.5%; Pred. No. 6.1e-207;					
	Matches 926; Conservative	0; Mismatches 86; Indels 0; Gaps 0;					
OY	1	AGACGCCGGTGAATGGCGACGCGCTGCAGACGAGTGCGCTTCATAAAGCTGTGCACC	60				
Db	997	AGAGAGACGGGTGACTCGGCGCGCTGCAGAGACGAGTGCGCTTCATAAAGCTGTGCACC	1056				
OY	61	CATGGATGAGACCAAGATCTTGATCTTAGGCCCACTGTAAATCCAGTAAGAATCTTGA	120				
Db	1057	CATGGATGAGACCAAGATCTTGATCTTAGGCCCTTCTAATAGATCCGATGGAAGACTCTGA	1116				
OY	121	CAACAGTCGCAATTATATGACAAGATTTAAAGACAGTGATGACTAGATGATCTGGCAGA	180				
Db	1117	CAACAGTCGCAATTATATGACAAGATTTAAAGACAAATGTGACTCTGGATGATCTGGCAGA	1176				
OY	181	CTTCTTTAAGAGTGTGGGGTTGTTAAGATGAACAAGAACTGGGGCAACCATGATCCA	240				
Db	1177	CTTCTTTAAGAGTGTGGGGTTGTTCAAGATGAACAAGAACTGGGCAACCATGATCCA	1233				
OY	241	CATCTACCTGACAAAGAACAGAAAAGCAAAGGCGATGCCACAGTGTCTTATGANA	300				
Db	1237	TATCTACTGGATTAAGAGACAGAAAAGCTTAAAGGGGACCCACAGTGTCTTATGANA	1296				
OY	301	CCCACCCACTGCCAAGGCTGCGGTGGAAATGTTTTGATGGGAAATTTTCAAGGAGCAA	360				
Db	1297	TCCACCACTCCAAAGGCTGCCGTGGAAATGTTTTGATGGGAAAGTTTTTCAAGGAGCAA	1356				
OY	361	ACTTAAATGTCCCTGTGCTGGAGAAAGCTCCCATGAAACGTAATGCGGGGTGCTGCC	420				
Db	1357	ACTTAAATGTCTTGTGCCGAAAGAACCTTCCATATGACACATGCGGGAGGCGATGCC	1411				
OY	421	AACCCGTGAGGCGAAGGCAATGCCACCAACAACCTCCGTGAGAGTCCAGAGGCCACGAGG	480				
Db	1417	AACCTGTATGAGGCGAGGGGATGCCACCAACAACCTCTGTGAGAGTCCGTGTGGCCAGAGG	1476				
OY	481	TCCTGGGGGACCACTGGGTCCGATGGAGGCGCGGTGAGAGAGATGAGAGGCTTCCCTTC	540				



Db 1477 CCTGAGAGACCAATGGTGGATGGAGGCGGTGGAGAGACAGAGGGGCTTCCCTCC 1536  
Qy 541 AAGAGACCCCGGGGTTCCGAGGAAACCCCTCTGAGAGAGAAACCTCCAGACCGAGC 600  
Db 1537 AAGAGGGCCCCGAGGCTCCAGAGGAAACCCCTCTGAGAGAGAAATGTCAGACCGAGC 1596  
Qy 601 TGGAGATTGACAGTGTCCCAATCCGGGTGTGGAAACCAAACTTCCCTGGAGAACAGA 660  
Db 1597 TGGAGATTGACAGTGTCCCAATCCGGGTGTGGAAACCAAACTTCCCTGGAGAACAGA 1656  
Qy 661 GTCCAAACAGTGTAAAGGCCCCAAAGCTGAAAGCTTCTCCCGCACCTTCCGCCCC 720  
Db 1657 ATGCAACCAAGTAAAGGCCCCCTTAAGCCGAGAGCTTCTCCCGCACCTTCCAGCTCC 1716  
Qy 721 GGGTGTGATGTGGACAGAGTGGCCCTGTGGCATCGGGAGAAAGAGTGGCTCAT 780  
Db 1717 GGGTGTGATGTGGACAGAGTGGCCCTGTGGCATCGAGAGAAAGAGAGACTCAT 1776  
Qy 781 GGAATCGTGTGTCCCGGTGGAATGTTCAAGAGTGGCCGTGTGGAGACAGAGTGGCTT 840  
Db 1777 GGAACGTGTGTCTTGGAGAAATGTTCAAGAGTGGAGAGAGTGGAGACAGAGAGCTT 1836  
Qy 841 CCGTGTGGCCGGGGCAATGACCGAGGTGGCTTGTGAGAGAAAGAGAGGCTCCCTGG 900  
Db 1837 CCGAGTGGCCGGGTGGAATGACCGAGGTGGCTTGTGAGAGAAAGAGAGTGGCTCC 1896  
Qy 901 GGGGCCCCCTTGAACCTTTGATGGAACAGATGGAGAGAAAGAGAGAGCTGGAGACC 960  
Db 1897 GGGGCCCCCTTGAACCTTTGATGGAACAGATGGAGAGAAAGAGAGAGCTGGAGACC 1956  
Qy 961 TGGAAAAATGATTAAGGCGAGCACCTCAGAGAGCGAGAGATGGCGCTTAC 1012  
Db 1957 TGGAAAAATGATTAAGGCGAGCACCTCAGAGAGCGAGAGATGGCGCTTAC 2008

## RESULT 14

LOCUS BX87336/c 1020 bp mRNA linear EST 27-APR-2004

DEFINITION BX87336 Homo sapiens NEUROBLASTOMA COT 50-NORMALIZED Homo sapiens

ACCESSION BX87336 CDS00004YE17 3-PRIME, mRNA sequence.

VERSION BX87336.1 GI:30455516

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

TITLE 1 (bases 1 to 1020)

JOURNAL Full-length cDNA libraries and normalization

COMMENT Unpublished (2001)

CONTACT: Genoscope

Genoscope - Centre National de Sequencage

2 rue Gaston Cremieux, CP 5706 - 91057 Evry cedex - FRANCE

Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr

1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime

end enriched, double-strand cDNA was digested with Not I and cloned

into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library

was normalized. Library was constructed by Life Technologies, a

division of Invitrogen. This sequence belongs to sequence cluster

10589.f

For more information about this cluster, see

http://www.genoscope.cns.fr/cdna?c=CS6AD0012D03NP1&c=10589.f.

location/Qualifiers

1.1020

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="CS0D004YE17"

/cissue="NEUROBLASTOMA COT 50-NORMALIZED"

/clone.lib="Homo sapiens NEUROBLASTOMA COT 50-NORMALIZED"

/note="1st strand cDNA was primed with a NotI-oligo(dT)."

## ORIGIN

primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized."

Query Match 82.5%; Score 834.8; DB 5; Length 1020;

Best Local Similarity 97.6%; Pred. No. 9.1e-197;

Matches 855; Conservative 11; Mismatches 8; Indels 2; Gaps 2;

Qy 137 GTACAAAGATTAAATGACAGTGTGACTCTGATGATCTGGACAGCTTTAAAGTGT 196  
Db 1020 KMCAAGATTAAATGACATTTACTCTGATGATCTGGACAGCTTTAAAGTGT 962  
Qy 197 GGGGTGTAAATGAAACAAGAGATCGGCAACCCATGATCCATCTTCTGACAG 256  
Db 961 GGGGTGTAAATGAAACAAGAGATCGGCAACCCATGATCCATCTTCTGACAG 902  
Qy 257 GAAACAGAAAGCCCAAGGCGATGCCAGTGTCTTATGAAGCCCACTGCCAG 316  
Db 901 GAAACAGAAAGCCCAAGGCGATGCCAGTGTCTTATGAAGCCCACTGCCAG 842  
Qy 317 GCTGCCGTGAAATGTTTATGAGAAAGATTTTCAAGGAGCAAACTTAAAGTCT 376  
Db 841 GCTGCCGTGAAATGTTTATGAGAAAGATTTTCAAGGAGCAAACTTAAAGTCT 782  
Qy 377 GCTGGAGAAAGCTTCCATGAAACAGTATCGGGGTGTCTTCCCACTGAGAGG 436  
Db 781 GCTGGAGAAAGCTTCCATGAAACAGTATCGGGGTGTCTTCCCACTGAGAGG 722  
Qy 437 GGGATGCAACACACCTCCGATGAGTCCAGAGGCGCCAGAGAGTCTTGGAGAG 496  
Db 721 GGGATGCAACACACCTCCGATGAGTCCAGAGGCGCCAGAGAGTCTTGGAGAG 662  
Qy 497 GGTGCAATGAGAGAGCCGTGAGAGATGAGAGAGCTTCCCTCAAGAGAGACC 556  
Db 661 GGTGCAATGAGAGAGCCGTGAGAGATGAGAGAGCTTCCCTCAAGAGAGACC 603  
Qy 603 GGTGCAATGAGAGAGCCGTGAGAGATGAGAGAGCTTCCCTCAAGAGAGACC 662  
Db 557 TCCGAGAGGAAACCCCTCTGAGAGAGAAAGTCCAGACCGAGCTGGAGTGT 616  
Qy 616 TCCGAGAGGAAACCCCTCTGAGAGAGAAAGTCCAGACCGAGCTGGAGTGT 676  
Db 602 TCCGAGAGGAAACCCCTCTGAGAGAGAAAGTCCAGACCGAGCTGGAGTGT 543  
Qy 543 TCCGAGAGGAAACCCCTCTGAGAGAGAAAGTCCAGACCGAGCTGGAGTGT 616  
Db 617 CCCAATCCGGGTGTGAGAAACCAAGACTTCCCTGAGAGAAAGTCCAGTGT 676  
Qy 676 CCCAATCCGGGTGTGAGAAACCAAGACTTCCCTGAGAGAAAGTCCAGTGT 736  
Db 542 CCCAATCCGGGTGTGAGAAACCAAGACTTCCCTGAGAGAAAGTCCAGTGT 483  
Qy 483 GCGCCAAAGCGCTGAAAGCTTCTCCGCGCACCTTCCGCGCGCGCGGTGTGAT 423  
Db 423 GCGCCAAAGCGCTGAAAGCTTCTCCGCGCACCTTCCGCGCGCGCGGTGTGAT 423  
Qy 737 AGAGTGGCCCTGTGAGATGCGGGAGAGAAAGTGTGCTCATGATGTGTGTGCC 796  
Db 422 AGAGTGGCCCTGTGAGATGCGGGAGAGAAAGTGTGCTCATGATGTGTGTGCC 363  
Qy 797 GGTGAAATGTTCAAGAGTGGCCCTGTGAGAGAAAGTGTGCTCATGATGTGTGCC 856  
Db 362 GGTGAAATGTTCAAGAGTGGCCCTGTGAGAGAAAGTGTGCTCATGATGTGTGCC 303  
Qy 857 ATGAGACGAGTGTGCTTGTGAGAGAAAGAGAGTGTGCTTGGGGGGCCCCCTG 916  
Db 302 ATGAGACGAGTGTGCTTGTGAGAGAAAGAGAGTGTGCTTGGGGGGCCCCCTG 243  
Qy 917 TTGATGAACAGATGAGAGAGAAAGAGAGAGTGTGAGAGAGCTTGAATAA 976  
Db 242 TTGATGAACAGATGAGAGAGAAAGAGAGAGAGTGTGAGAGAGCTTGAATAA 183  
Qy 977 GGGAGACACCTTGAAGAGAGAGAGAGAGTGTGAGAGAGCTTGAATAA 1012  
Db 182 GGGAGACACCTTGAAGAGAGAGAGAGAGAGTGTGAGAGAGCTTGAATAA 147

## RESULT 15

LOCUS AK049743 2269 bp mRNA linear HTC 03-APR-2004

DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM
Mus musculus 12 days embryo spinal cord cDNA, RIKEN full-length enriched library, clone:C530046A18 product:Ewing sarcoma histology, full insert sequence.	AK049743	AK049743.1	GI:26093614	HTC; CAP trapper.	Mus musculus (house mouse)
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
MEDLINE					
PUBMED					
REFERENCE					
AUTHORS					
1	Carninci, P. and Hayashizaki, Y.				
2	High-efficiency full-length cDNA cloning				
3	Meth. Enzymol. 303, 19-44 (1999)				
4	99279253				
5	10349636				
6	Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.				
7	Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes				
8	Genome Res. 10 (10), 1617-1630 (2000)				
9	204993174				
10	11042159				
11	Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Kono, H., Akiyama, J., Nishi, K., Kitenaka, T., Tashiro, H., Itoh, M., Suni, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kasai, Y., Fujikawa, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watabiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.				
12	RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer				
13	Genome Res. 10 (11), 1757-1771 (2000)				
14	11076861				
15	The RIKEN Genome Exploration Research Group Phase II Team and the PANTOM Consortium.				
16	Functional annotation of a full-length mouse cDNA collection				
17	Nature 409, 685-690 (2001)				
18	5				
19	The PANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.				
20	Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs				
21	Nature 420, 563-573 (2002)				
22	6 (bases 1 to 2269)				
23	Adachi, J., Aizawa, K., Akimura, T., Aikawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W., Hayashizaki, K., Hayatsu, N., Hizemoto, K., Hirooka, T., Hirozane, T., Horii, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Kono, H., Kouda, M., Koya, S., Kurihar, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Nunazaki, R., Ohno, M., Ohsato, N., Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shingawa, A., Shiraki, T., Soabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toyo, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.				
24	Muramatsu, M. and Hayashizaki, Y.				
25	Direct Submission				
26	Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration and Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:genome-research@riken.jp, URL: http://genome.gsc.riken.jp/, Tel:81-45-503-9222, Fax:81-45-503-9216)				
27	CDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in Riken Division of Experimental Animal Research in Riken contributed to				

FEATURES		source	
1. .2269		/organism="Mus musculus"	
		/mol_type="mRNA"	
		/strain="C57BL/6J"	
		/db_xref="RANTOM:DB:C530046A18"	
		/db_xref="taxon:10090"	
		/clone="C530046A18"	
		/tissue_type="spinal cord"	
		/clone_lib="RIKEN full-length enriched mouse cDNA library"	
		/dev_stage="12 days embryo"	
		48..1903	
misc_feature		/note="Ewing sarcoma homolog (MGI:99960, GB NM_007968, evidence: BLASTN, 9%, match=2172)	
		putative"	
		2251..2256	
polya_signal		/note="putative"	
		2269	
polya_site		/note="putative"	
ORIGIN			
Query Match		78.3%; Score 792.8; DB 3; Length 2269;	
Best Local Similarity		90.2%; Pred. No.3.1e-186;	
Matches 859; Conservative		0; Mismatches 92; Indels 1; Gaps 1;	
QY	61	CATGATGAAGACACAGATCTTGATCTTAGGCCCACTGTGTGATCCAGATGAAGACTTGA	120
DB	950	CAGGGGAAGGGGGATTGATCGTGAGGCATGACAGAGATCCCATGAAAGACTTGA	1009
QY	121	CAAGAGCAATTATATACAAAGATTAATAATGACAGTGTACTTAGATGATCTGCGAGA	180
DB	1010	CAACAGCAATTTATGTGCAAGGATTAATGACATGTACTGTGATGTCTGCGAGA	1069
QY	181	CTTCTTTAAGCAGTGTGGGTTGTTAAGATGAA CAAAGAACTGGGCAACCCATGATCA	240
DB	1070	CTTCTTTAAGCAGTGTGGGTTGTTAAGATGAA CAAAGAACTGGGCAACCCATGATCA	1129
QY	241	CATCTACCTGGACAAAGAA CAGGAAAGCCCAAGCGCATGCCACATGTCTTATGAAGA	300
DB	1130	TATCTACCTGGATTAAGAGACAGGAAAGCCTAAAGGGGAAGCCACATGTCTTATGAAGA	1189
QY	301	CCGACCCACCTGCCAAGGCTGCCGTGGAGTATGATGGGAAAGATTTTCAAGGGACAA	360
DB	1190	TCACCAACTGCAAAAGGCTGCCGTGGAGTATGATGGGAAAGATTTTCAAGGACAA	1249
QY	361	ACTTAAAGTCTCCCTTCTCGGAAGAAAGCTTCAATGAACATATGCGGGGTGTCTGCC	420
DB	1250	ACTTAAAGTGTCTTCTCCGGAAGAAAGCTTCAATGAACATATGCGGGGTGTCTGCC	1309
QY	421	ACCCCTGAGAGGACAGAGCATGCCACCACTCCGTGGAGGTTCAGAGGCCAGAGAG	480
DB	1310	ACCTCGTGAAGGACAGGAGATGCCACCACTTCGTGAGGTCTCTGTGCCCTGAGAGG	1369
QY	481	TCCTGGGGGACCCATGTGGGTGCGATGGGAGGCCGTGAGAGATGTAGAGAGGCTTCCCTCC	540
DB	1370	CCCTGAGAGACCCATGTGGGTGCGATGGGAGGCCGTGAGAGAGATGTAGAGAGGCTTCCCTCC	1429
QY	541	AAGAGAACCCCGGGGTTCCGAGGGAAAGCTTCTGAGAGAGAAAGCTTCACGACCGAGC	600
DB	1430	AAGAGAGGCCCGGAGGCTTCAGAGGAAAGCTTCTGAGAGAGAAAGTTCACGACCGAGC	1489
QY	601	TGAGAGCTGGCAGTGTCCCAATCCGGGTGTGGAAACCAAGATTTGCCCTGGAGAAACAGA	660
DB	1490	TGAGAGCTGGCAGTGTCCCAATCCGGGTGTGGAAACCAAGATTTGCCCTGGAGAAACAGA	1549
QY	661	GTGACAACAGGTGAAGGCCCAAGAGCTGGAAGCTTCTCCCGCACCTTTCCGCGCCC	720
DB	1550	ATGACAACAGGTGAAGGCCCTGAAGGCCGAGGCTTCTCCCGCACCTTTCCACCTCC	1609

```
QY 721 GGGTGGTGAATCGTGGACAGAGTGGCCCTGGTGGCAATCGGGGAGAAAGAGTGGCTCAT 780
Db 1610 GGGTGGTGAATCGTGGACAGAGTGGCCCTGGTGGCAATCGAGAGAAAGAGGAGTCAAT 1669
QY 781 GGAATCGTGGTGGTCCCGGTGAAATGTTCAAGAGTGGCCGTGGTGGAGACAGAGTGGCTT 840
Db 1670 GGAACGTGGTGGTCTCTGGAGAAATGTTCAAGAGTGGCAAGAGTGGAGACAGAGAGGCTT 1729
QY 841 CCGTGGTGGCCGGGGCAATGGAACGAGGTGGCTTTGGTGGAGAAAGACGAGGTGGCCCTGG 900
Db 1730 CCGAGGTGGCC-GGGAATGGACCGAGGTGGCTTTGGTGGAGAAAGACGAGGTGGTCTGG 1788
QY 901 GGGGCCCCCTGGACCTTTGAATGGAACAGATGGAGGAAAGAGAGAGAGTGGAGGACC 960
Db 1789 GGGGCTCTCTGGACCTTTAAATGGAAACAGATGGAGGAAAGAGAGCGGACGTGGAGGACC 1848
QY 961 TGGAAAAAATGATPAAAGCGAGCAACGTCAGAGACGAGAGATGGCCCTTAC 1012
Db 1849 TGGGAAAAATGATPAAAGCGAGCAACCTCTCAGGAAACGAGACCGGCCCTTAC 1900
```

Search completed: February 20, 2005, 16:17:55  
Job time : 2994.19 secs

**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Comugen Ltd.

# OM protein - protein search, using sw model

Run on: February 18, 2005, 14:30:59, Search time 172 Seconds  
(without alignments)  
1475.086 Million cell updates/sec

Title: US-10-791-017A-2

Perfect score: 3633  
Sequence: 1 MASTDSTYSQAAAQGYSA.....GGPGKDKGKHROERDRPY 656

Scoring table: BLOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

A\_Geneseq\_16Dec04:\*  
1: geneseqp1980s:\*  
2: geneseqp1990s:\*  
3: geneseqp2000s:\*  
4: geneseqp2001s:\*  
5: geneseqp2002s:\*  
6: geneseqp2003as:\*  
7: geneseqp2003bs:\*  
8: geneseqp2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysts of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3633	100.0	656	8	ADSL6278 Human Ew1
2	3633	100.0	656	8	ADSL6278 Human Ew1
3	3624	99.8	656	2	AAK44555 Human EWS
4	3611	99.4	656	2	AAW33813 Human EWS
5	3600.5	99.1	655	4	ABG06460 Novel hum
6	3571.5	98.3	655	5	ABBS7126 Mouse lbc
7	3277	90.2	600	6	ADA55145 Human pro
8	3210	88.4	604	6	ADRO9402 Human pro
9	3178.5	87.5	603	8	ADP56334 Human PRO
10	1639	45.1	361	8	ABM80197 Tumour-as
11	1429	39.3	476	5	ABG95068 Human tra
12	1420	39.1	362	2	AAW34972 Chimeric
13	1420	39.1	365	2	AAW34971 Chimeric
14	1250	35.5	518	8	AD126113 Human pro
15	1282	35.3	526	4	AAW78355 Human pro
16	1282	35.3	526	5	ABG95081 Human tra
17	1282	35.3	526	8	AD126117 Human pro
18	1282	35.3	526	8	AD126117 Human pro
19	1282	35.3	526	8	AD126117 Human pro
20	1281.5	35.3	525	8	AD126117 Human pro
21	1281.5	35.3	525	8	AD126117 Human pro
22	1273	35.0	536	4	AAW78356 Human NF-
23	1263	34.8	569	4	AAW79339 Human pro
24	1263	34.8	569	4	AAW79340 Human pro
25	1009	27.8	592	6	ABOS3050 Human put

26	1009	27.8	592	8	AD126209	AD126209 Human pro
27	1003.5	27.6	589	8	AD126207	AD126207 Human pro
28	799.5	22.0	260	7	ADJ68310	ADJ68310 Human hea
29	717.5	19.7	399	4	ABBE6010	ABBE6010 Drosophi1
30	623	17.1	156	3	AAK44066	AAK44066 Human can
31	621	17.1	280	8	AD126115	AD126115 Human pro
32	531.5	14.6	266	7	AD122735	AD122735 Human dis
33	484	13.3	295	2	AAK44579	AAK44579 EWS gene/
34	475	13.1	83	4	AAW21854	AAW21854 Peptide #
35	475	13.1	83	4	ABBA44223	ABBA44223 Peptide #
36	475	13.1	83	4	AAW38178	AAW38178 Peptide #
37	475	13.1	83	4	ABBS7099	ABBS7099 Protein #
38	475	13.1	83	4	AAW77960	AAW77960 Human bon
39	475	13.1	83	4	AAW65255	AAW65255 Human bra
40	475	13.1	83	4	ABG59596	ABG59596 Human liv
41	475	13.1	83	5	ABG46973	ABG46973 Human pep
42	472.5	13.0	462	5	ABG95083	ABG95083 Human tra
43	470.5	13.0	462	8	ADK77159	ADK77159 Human pro
44	400.5	11.0	1466	4	ABBS0291	ABBS0291 Collagen
45	400.5	11.0	1466	5	ABBS0747	ABBS0747 Human Tum

## ALIGNMENTS

RESULT 1  
ADSL6278 standard; protein; 656 AA.  
XX  
AC ADSL6278;  
XX  
DT 18-NOV-2004 (first entry)  
XX  
DE Human Ewing sarcoma protein Segid 2.  
XX  
KW human; Ewing sarcoma; EWS; prostatic cancer; alopecia; acne;  
XX hypogonadism; androgen-resistance syndrome; testicular feminisation.  
XX  
OS Homo sapiens.  
XX  
PN EP1455190-AA.  
XX  
PD 08-SEP-2004.  
XX  
PF 16-FEB-2004; 2004EP-00003422.  
XX  
PR 04-MAR-2003; 2003DE-01009280.  
XX 25-APR-2003; 2003US-0465692P.  
XX  
PA (SCHD) SCHERING AG.  
XX  
PI Obendorf M, Wolf S;  
XX  
XX WPI, 2004-627861/61.  
XX N-PSDB; ADSL6277.  
XX  
XX Determining the hormonal effects of substances, used to identify  
XX pharmaceuticals, e.g. for treatment of androgen receptor dysfunction,  
XX from modulating interaction between nuclear receptors and Ewing sarcoma  
XX protein.  
XX  
XX Disclosure; SEQ ID NO 2; 30pp; German.  
XX  
XX This invention relates to a novel modulators that alter the interaction  
XX between the Ewing sarcoma protein (EWS) and its nuclear receptor, as well  
XX as the screening method thereof. Specifically, it refers to determining  
XX and identifying a hormonal effect brought about by test compounds that  
XX modulate either the binding of EWS to the nuclear receptor or the ligand-  
XX induced activity of this receptor. The present invention describes the  
XX nuclear receptors as including oestrogen, progesterone, thyroid hormone,  
XX vitamin D, and retinoic acid receptors, most preferably they are androgen  
XX receptors. Accordingly, these modulators may be used in the development  
XX of pharmaceutical compositions that can diagnose and be used to treat

CC diseases associated with receptor dysfunction such as prostatic cancer,  
 CC alopecia, acne, hypogonadism and androgen-resistance syndrome e.g.  
 CC testicular feminisation. This method provides reliable, sensitive,  
 CC simple, inexpensive and rapid assessment of the hormonal effects of these  
 CC test compounds. This polypeptide sequence is the human Ewing sarcoma  
 CC protein of the invention.

XX Sequence 656 AA;

Query Match 100.0%; Score 3633; DB 8; Length 656;  
 Best Local Similarity 100.0%; Pred. No. 1.4e-216;  
 Matches 656; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MASTDVSITYSQAAGQGYSAITAPOTGVAQTTOAGGQSYGYGQPTDVSITQAQTAT 60  
 DB 1 MASTDVSITYSQAAGQGYSAITAPOTGVAQTTOAGGQSYGYGQPTDVSITQAQTAT 60  
 QY YGQTAATATSGQPTGTTTPTAPQAYSQPVQGYGTGAYDTTATVTTTQASVAAQSAVGT 120  
 DB 61 YGQTAATATSGQPTGTTTPTAPQAYSQPVQGYGTGAYDTTATVTTTQASVAAQSAVGT 120  
 QY 121 QPAYPAYGQOPATATPAPTRPQDGNKPTETSPQSSSTGYNPSLQYGSNTSYQVPGSY 180  
 DB 121 QPAYPAYGQOPATATPAPTRPQDGNKPTETSPQSSSTGYNPSLQYGSNTSYQVPGSY 180  
 QY 181 MQPVTAPSPYPTSYSTOPTSYDOSSYQONTYGGPSSYGGQSSYGGQSPPTSY 240  
 DB 181 MQPVTAPSPYPTSYSTOPTSYDOSSYQONTYGGPSSYGGQSSYGGQSPPTSY 240  
 QY 241 PPQTGSYSQAPSGYSQSSSYGGQSSFRQDHPSSMGVYQGESGSGFSGPGENRMSGPDNR 300  
 DB 241 PPQTGSYSQAPSGYSQSSSYGGQSSFRQDHPSSMGVYQGESGSGFSGPGENRMSGPDNR 300  
 QY 301 GRGGRGFRGSGMRGGRGGRGSGAGRGGRGFKPCGPMDEGPDLDLGPVPDDESDN 360  
 DB 301 GRGGRGFRGSGMRGGRGGRGSGAGRGGRGFKPCGPMDEGPDLDLGPVPDDESDN 360  
 QY 361 SATYVOGLNLSVTLDDLADFPKCGVYVKNMKRTGQPMIHIYLDKRGKRPDGAATVSYEDP 420  
 DB 361 SATYVOGLNLSVTLDDLADFPKCGVYVKNMKRTGQPMIHIYLDKRGKRPDGAATVSYEDP 420  
 QY 421 PTKAAAVFWFDKDFQGSXKLKVLARKKPPMNSMRGGLPREGRGMPPLRGPGFGPG 480  
 DB 421 PTKAAAVFWFDKDFQGSXKLKVLARKKPPMNSMRGGLPREGRGMPPLRGPGFGPG 480  
 QY 481 GGPWGRMGSGRGGFRPPRGPRGRGNPSGGGVQHRADWQCPNPGCANNFAMTEC 540  
 DB 481 GGPWGRMGSGRGGFRPPRGPRGRGNPSGGGVQHRADWQCPNPGCANNFAMTEC 540  
 QY 541 NCKAPRBEGLPPLPPPGDGRGGRGPGMGRGGGLMDRGGCGMFRGGRGGDRGGR 600  
 DB 541 NCKAPRBEGLPPLPPPGDGRGGRGPGMGRGGGLMDRGGCGMFRGGRGGDRGGR 600  
 QY 601 GGRGMDRGFGGGRGGRGPGPLPMEQMGRRGGRGPGKMDGKHRRQRRDPY 656  
 DB 601 GGRGMDRGFGGGRGGRGPGPLPMEQMGRRGGRGPGKMDGKHRRQRRDPY 656

RESULT 2  
 ABM82330  
 ID ABM82330 standard; protein; 656 AA.

XX ABM82330;  
 DT 18-NOV-2004 (first entry)  
 XX Tumour-associated antigenic target (TAT) polypeptide PRO58232, SEQ:5990.  
 XX Tumour-associated antigenic target; TAT; human; overexpression; cancer;  
 KM tumour; diagnosis; cell proliferative disorder; breast cancer;  
 KM colorectal cancer; lung cancer; ovarian cancer; liver cancer;  
 KM central nervous system cancer; bladder cancer; pancreatic cancer;  
 KM cervical cancer; melanoma; leukaemia; hybridisation probe;

KM chromosome identification; chromosome mapping; gene mapping;  
 KM gene therapy; cytostatic.

XX Homo sapiens.

XX WO2004030615-A2.

XX 15-APR-2004.

XX 29-SEP-2003; 2003WO-US028547.

XX 02-OCT-2002; 2002US-041971P.

XX (GETH ) GENENTECH INC.

XX Wu TD, Zhang Z, Zhou Y;

XX WPI; 2004-347921/32.

XX N-PSDB; ACN40903.

PT New tumor-associated antigenic target polypeptides and nucleic acids,  
 PT useful in preparing a medicament for treating or detecting a  
 PT proliferative disorder, e.g. breast, lung, colorectal, ovarian or  
 PT prostate cancer or tumor.

PS Claim 12; SEQ ID NO 5990; 7273pp; English.

CC The invention relates to human tumour-associated antigenic target (TAT)  
 CC polypeptides, and their related nucleic acids. The TAT polypeptides are  
 CC overexpressed in cancer tissues compared to normal tissues, and may thus  
 CC serve as effective targets for the diagnosis and treatment of cancer in  
 CC mammals. The invention also relates to nucleic acid and polypeptide  
 CC sequences at least 80% identical to the TAT nucleic acids and  
 CC polypeptides, expression vectors and host cells comprising a TAT nucleic  
 CC acid; an antibody specific for a TAT polypeptide; a peptide or organic  
 CC molecule which binds to a TAT polypeptide; fusion proteins comprising a  
 CC TAT polypeptide; and methods and compositions for the treatment or  
 CC diagnosis of cancer in mammals. TAT polypeptides, nucleic acids,  
 CC antibodies, antagonists, binding molecules and compositions are useful  
 CC for diagnosing or treating a cell proliferative disorder associated with  
 CC increased TAT expression, particularly cancers such as breast cancer,  
 CC colorectal cancer, lung cancer, ovarian cancer, liver cancer, bladder  
 CC cancer, pancreatic cancer, cervical cancer, cancers of the central  
 CC nervous system, melanoma and leukaemia. TAT nucleic acids may further be  
 CC used as hybridisation probes, in chromosome and gene mapping, in  
 CC chromosome identification and in gene therapy. The present sequence  
 CC represents a TAT polypeptide of the invention

XX Sequence 656 AA;

Query Match 100.0%; Score 3633; DB 8; Length 656;  
 Best Local Similarity 100.0%; Pred. No. 1.4e-216;  
 Matches 656; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MASTDVSITYSQAAGQGYSAITAPOTGVAQTTOAGGQSYGYGQPTDVSITQAQTAT 60  
 DB 1 MASTDVSITYSQAAGQGYSAITAPOTGVAQTTOAGGQSYGYGQPTDVSITQAQTAT 60  
 QY 61 YGQTAATATSGQPTGTTTPTAPQAYSQPVQGYGTGAYDTTATVTTTQASVAAQSAVGT 120  
 DB 61 YGQTAATATSGQPTGTTTPTAPQAYSQPVQGYGTGAYDTTATVTTTQASVAAQSAVGT 120  
 QY 121 QPAYPAYGQOPATATPAPTRPQDGNKPTETSPQSSSTGYNPSLQYGSNTSYQVPGSY 180  
 DB 121 QPAYPAYGQOPATATPAPTRPQDGNKPTETSPQSSSTGYNPSLQYGSNTSYQVPGSY 180  
 QY 181 MQPVTAPSPYPTSYSTOPTSYDOSSYQONTYGGPSSYGGQSSYGGQSPPTSY 240  
 DB 181 MQPVTAPSPYPTSYSTOPTSYDOSSYQONTYGGPSSYGGQSSYGGQSPPTSY 240  
 QY 241 PPQTGSYSQAPSGYSQSSSYGGQSSFRQDHPSSMGVYQGESGSGFSGPGENRMSGPDNR 300  
 DB 241 PPQTGSYSQAPSGYSQSSSYGGQSSFRQDHPSSMGVYQGESGSGFSGPGENRMSGPDNR 300



```

QY 301 GGRGGFDRGMSRGGGSGAGRGFTNKGPMDEGPDLDLGPVPDPEDSDN 360
DB 301 GGRGGFDRGMSRGGGSGAGRGFTNKGPMDEGPDLDLGPVPDPEDSDN 360
QY 361 SAITYGGLSDSVTLDDLADFFKCCGVVKNKRTGQPMIHLYLDEKTKPKGDAIVSYEDP 420
DB 361 SAITYGGLSDSVTLDDLADFFKCCGVVKNKRTGQPMIHLYLDEKTKPKGDAIVSYEDP 420
QY 421 PTKAAVWFVDFDGDPOGSKLKVSLARKKPPMNSMRGGLPPREGRGMPPLRGPGGPGP 480
DB 421 PTKAAVWFVDFDGDPOGSKLKVSLARKKPPMNSMRGGLPPREGRGMPPLRGPGGPGP 480
QY 481 GGPMMGMRGGDRGGFPPRPGSRGNPSGGANVOHRAGDMOCNPGCCGNQNFAMRTEC 540
DB 481 GGPMMGMRGGDRGGFPPRPGSRGNPSGGANVOHRAGDMOCNPGCCGNQNFAMRTEC 540
QY 541 NCKAKRPPGFLPPPPPGGDRGGGPGGMRGGGLMDRGPGGMFRGGGDRGGFR 600
DB 541 NCKAKRPPGFLPPPPPGGDRGGGPGGMRGGGLMDRGPGGMFRGGGDRGGFR 600
QY 601 GGRGMDRGFGGRRGGPGPPGLMEOMGRRGRRGPGKMDKGHRQERRDRPY 656
DB 601 GGRGMDRGFGGRRGGPGPPGLMEOMGRRGRRGPGKMDKGHRQERRDRPY 656

```

## RESULT 3

AAR44555

ID AAR44555 standard; protein; 656 AA.

XX AAR44555;

DT 25-MAR-2003 (revised)  
 DT 26-MAY-1994 (first entry)

DE Human EWS protein deduced from foetal brain CDNA clone B1A5.

KW chromosomal translocation; chimeric; chimaeric; Ewing sarcoma; Ews gene;  
 KW malignant melanoma; human chromosome 11;  
 KW primitive peripheral neuroectodermal tumour; human chromosome 22;  
 KW hum-fl1-1.

OS Homo sapiens.

XX Key Location/Qualifiers  
 FT M18c-difference 569  
 FT /note= "corresponds to GGT codon"

XX MO9323549-A2.

XX 25-NOV-1993.

XX 19-MAY-1993; 93WO-FR000494.

XX 20-MAY-1992; 92FR-00006123.

XX (CNRS ) CNRS CENT NAT RECH SCI.

XX Aurias A, Delattre O, Desmaza C, Melot T, Peter M, Plougastel B;

XX Thomas G, Zucman J;

XX WPI; 1993-386580/48.

XX N-PSDB; AAO50643.

XX New nucleic acid of EWS gene and its hybrid(s) - contg. gene sequence

XX involved in chromosomal translocation, also derived mRNA, probes, fusion

XX proteins etc., for diagnosis and treatment of Ewing sarcoma and melanoma.

XX Disclosure; Fig 6; 123p; French.

XX The probes 22R3 and 22R12 were used to screen a human foetal brain CDNA

XX library (Stratagene cat. # 936206). The clone B1A5 was identified and

XX sequenced. It represents the entire coding region and 3'-UTR of the Ews

CC gene. (updated on 25-MAR-2003 to correct PN field.)

XX SQ Sequence 656 AA;

Query Match 99.8%; Score 3624; DB 2; Length 656;  
 Best Local Similarity 99.8%; Pred. No. 4.9e-216;  
 Matches 655; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

QY 1 MASTDYSTYSQAQAAGGSAATTAQPTQGYAQTTOAYGQGSYGTGPTDVSYTOAQTAT 60
DB 1 MASTDYSTYSQAQAAGGSAATTAQPTQGYAQTTOAYGQGSYGTGPTDVSYTOAQTAT 60
QY 61 YGCTAATSYGQPTPTTTPAQAASOPVQGGTAATTTTATTTTAAQSAAYGT 120
DB 61 YGCTAATSYGQPTPTTTPAQAASOPVQGGTAATTTTATTTTAAQSAAYGT 120
QY 121 QPAYPAYGQOPATAATPRTPODGNKPTETSPQOSTGGYNPSLGYQSNYSYQVPGSY 180
DB 121 QPAYPAYGQOPATAATPRTPODGNKPTETSPQOSTGGYNPSLGYQSNYSYQVPGSY 180
QY 181 MQPVTAPSYPTSYSTQPTSYDQSSYQONTYQPPSSYGQSSYQOQPTSY 240
DB 181 MQPVTAPSYPTSYSTQPTSYDQSSYQONTYQPPSSYGQSSYQOQPTSY 240
QY 241 PPGTGSYSQAPSGYSQOQSSYQOQSSYQOQSSYQOQSSYQOQSSYQOQSSYQOQSSY 300
DB 241 PPGTGSYSQAPSGYSQOQSSYQOQSSYQOQSSYQOQSSYQOQSSYQOQSSYQOQSSY 300
QY 301 GGRGGFDRGMSRGGGSGAGRGFTNKGPMDEGPDLDLGPVPDPEDSDN 360
DB 301 GGRGGFDRGMSRGGGSGAGRGFTNKGPMDEGPDLDLGPVPDPEDSDN 360
QY 361 SAITYGGLSDSVTLDDLADFFKCCGVVKNKRTGQPMIHLYLDEKTKPKGDAIVSYEDP 420
DB 361 SAITYGGLSDSVTLDDLADFFKCCGVVKNKRTGQPMIHLYLDEKTKPKGDAIVSYEDP 420
QY 421 PTKAAVWFVDFDGDPOGSKLKVSLARKKPPMNSMRGGLPPREGRGMPPLRGPGGPGP 480
DB 421 PTKAAVWFVDFDGDPOGSKLKVSLARKKPPMNSMRGGLPPREGRGMPPLRGPGGPGP 480
QY 481 GGPMMGMRGGDRGGFPPRPGSRGNPSGGANVOHRAGDMOCNPGCCGNQNFAMRTEC 540
DB 481 GGPMMGMRGGDRGGFPPRPGSRGNPSGGANVOHRAGDMOCNPGCCGNQNFAMRTEC 540
QY 541 NCKAKRPPGFLPPPPPGGDRGGGPGGMRGGGLMDRGPGGMFRGGGDRGGFR 600
DB 541 NCKAKRPPGFLPPPPPGGDRGGGPGGMRGGGLMDRGPGGMFRGGGDRGGFR 600
QY 601 GGRGMDRGFGGRRGGPGPPGLMEOMGRRGRRGPGKMDKGHRQERRDRPY 656
DB 601 GGRGMDRGFGGRRGGPGPPGLMEOMGRRGRRGPGKMDKGHRQERRDRPY 656

```

## RESULT 4

AAM33813

ID AAM33813 standard; protein; 656 AA.

XX AAM33813;

XX 06-JUL-1998 (first entry)

XX Human EWS protein.

XX Tat stimulatory factor; Tat-SF1; transcriptional activity factor; HIV-1;

XX infection; gene therapy; EWS.

XX Homo sapiens.

XX Key Location/Qualifiers

XX Region 30..44  
 FT /note= "imperfect repeat peptide showing homology to Tat-SF1"  
 FT Region 209..236

FT	XX	/note= "imperfect repeat peptide showing homology to Tat
FT	XX	SPI"
FN	XX	
XX	XX	W09800695-A2.
XX	XX	
PD	XX	08-JAN-1998.
XX	XX	
PE	XX	03-JUL-1997; 97WO-US011713.
XX	XX	
PR	XX	03-JUL-1996; 96US-0021218P.
PR	XX	13-DEC-1996; 96US-0033152P.
XX	XX	
PA	XX	(MASI ) MASSACHUSETTS INST TECHNOLOGY.
XX	XX	
PI	XX	Sharp PA, Zhou Q;
XX	XX	
DR	XX	WPI; 1998-087086/08.
XX	XX	
PT	XX	Nucleic acid encoding Tat stimulating factor protein and related
PT	XX	transformed cells - proteins and binding agents, used to treat human
PT	XX	immunodeficiency virus infection.
XX	XX	
PS	XX	Example 6; Page 47-51; 68pp; English.
XX	XX	
CC	CC	EMS is a member of a novel family of putative transcription factors that
CC	CC	have RNA recognition motifs and which are frequently associated with many
CC	CC	types of sarcomas. It shows homology to a novel human transcriptional
CC	CC	activator factor, Tat-stimulatory factor (Tat-SF, see AAW3811), that is
CC	CC	involved in the regulation of transcriptional elongation of HIV-1 by Tat,
CC	CC	is essential for Tat trans-activation and is a substrate of an associated
CC	CC	cellular kinase. Tat-SF can be used to screen for binding agents useful
CC	CC	in the treatment of HIV infection
XX	XX	
SO	XX	Sequence 656 AA;

Query Match	Similarity	99.4%	Score	3611	DB 2	Length	656
Best Local	Similarity	99.7%	Pred. No.	3,1e-215			
Matches	654	Conservative	0	Mismatches	2	Indels	0
QY	1	MASTDVSYTSGAAAQGGYSAYTAQPTGQYAGTTQA YGQGSYGTGYPGPTDVSYTAQTTAT	60				
Db	1	MASTDVSYTSGAAAQGGYSAYTAQPTGQYAGTTQA YGQGSYGTGYPGPTDVSYTAQTTAT	60				
QY	61	YGGTAAYATSYGQPPRYGTTPTAPAPQASQPVGSGVGAADTTATTTTAASAAQASAGT	120				
Db	61	YGGTAAYATSYGQPPRYGTTPTAPAPQASQPVGSGVGAADTTATTTTAASAAQASAGT	120				
QY	121	QPAYPAYGQGPAAATAPTRPQDGNKRTETESQPSSTGYNQPSLIGQGSNYSYPCVPGSYP	180				
Db	121	QPAYPAYGQGPAAATAPTRPQDGNKRTETESQPSSTGYNQPSLIGQGSNYSYPCVPGSYP	180				
QY	181	MQPVTAPPSYPTTSYSTOPTSYDSSYSQDNTTYQPSSTYGGQSSSYGQGPPTSY	240				
Db	181	MQPVTAPPSYPTTSYSTOPTSYDSSYSQDNTTYQPSSTYGGQSSSYGQGPPTSY	240				
QY	241	PPQTSYSGAASQSYGQSSSYGQSSFPQDHPSSMGVYGQSSGSGGGRNSMGSPNNR	300				
Db	241	PPQTSYSGAASQSYGQSSSYGQSSFPQDHPSSMGVYGQSSGSGGGRNSMGSPNNR	300				
QY	301	GRGRGFPDRGMSRGRGGRGGRGMSAGBERGGFNKPGPMDEGPDLDLGPVYDDESDN	360				
Db	301	GRGRGFPDRGMSRGRGGRGGRGMSAGBERGGFNKPGPMDEGPDLDLGPVYDDESDN	360				
QY	361	SAIYVQGLNDSTLTLDLDLDFKQCGGVMMNRKTATGQPMHIYLDKRTGKPKGATYSYDP	420				
Db	361	SAIYVQGLNDSTLTLDLDLDFKQCGGVMMNRKTATGQPMHIYLDKRTGKPKGATYSYDP	420				
QY	421	PTAKAAVEMPDGKDFQGSKTLKYSLARKKXPMNNSMGGLPFRGRGMPPLRLRGQPGSGP	480				
Db	421	PTAKAAVEMPDGKDFQGSKTLKYSLARKKXPMNNSMGGLPFRGRGMPPLRLRGQPGSGP	480				
QY	481	CGPMGRMGGRGGRGGRGPPRGRGRGNPSSGGGANTQHPAGDMQCNPFGCANONFMRTEC	540				
Db	481	CGPMGRMGGRGGRGGRGPPRGRGRGNPSSGGGANTQHPAGDMQCNPFGCANONFMRTEC	540				

[illegible]

RESULT 5  
 ABG06460  
 ID ABG06460 standard; protein; 665 AA.  
 AC ABG06460;  
 AD  
 AE  
 AF  
 AG  
 AH  
 AI  
 AJ  
 AK  
 AL  
 AM  
 AN  
 AO  
 AP  
 AQ  
 AR  
 AS  
 AT  
 AU  
 AV  
 AW  
 AX  
 AY  
 AZ  
 BA  
 BB  
 BC  
 BD  
 BE  
 BF  
 BG  
 BH  
 BI  
 BJ  
 BK  
 BL  
 BM  
 BN  
 BO  
 BP  
 BQ  
 BR  
 BS  
 BT  
 BU  
 BV  
 BW  
 BX  
 BY  
 BZ  
 CA  
 CB  
 CC  
 CD  
 CE  
 CF  
 CG  
 CH  
 CI  
 CJ  
 CK  
 CL  
 CM  
 CN  
 CO  
 CP  
 CQ  
 CR  
 CS  
 CT  
 CU  
 CV  
 CW  
 CX  
 CY  
 CZ  
 DA  
 DB  
 DC  
 DD  
 DE  
 DF  
 DG  
 DH  
 DI  
 DJ  
 DK  
 DL  
 DM  
 DN  
 DO  
 DP  
 DQ  
 DR  
 DS  
 DT  
 DU  
 DV  
 DW  
 DX  
 DY  
 DZ  
 EA  
 EB  
 EC  
 ED  
 EE  
 EF  
 EG  
 EH  
 EI  
 EJ  
 EK  
 EL  
 EM  
 EN  
 EO  
 EP  
 EQ  
 ER  
 ES  
 ET  
 EU  
 EV  
 EW  
 EX  
 EY  
 EZ  
 FA  
 FB  
 FC  
 FD  
 FE  
 FF  
 FG  
 FH  
 FI  
 FJ  
 FK  
 FL  
 FM  
 FN  
 FO  
 FP  
 FQ  
 FR  
 FS  
 FT  
 FU  
 FV  
 FW  
 FX  
 FY  
 FZ  
 GA  
 GB  
 GC  
 GD  
 GE  
 GF  
 GG  
 GH  
 GI  
 GJ  
 GK  
 GL  
 GM  
 GN  
 GO  
 GP  
 GQ  
 GR  
 GS  
 GT  
 GU  
 GV  
 GW  
 GX  
 GY  
 GZ  
 HA  
 HB  
 HC  
 HD  
 HE  
 HF  
 HG  
 HH  
 HI  
 HJ  
 HK  
 HL  
 HM  
 HN  
 HO  
 HP  
 HQ  
 HR  
 HS  
 HT  
 HU  
 HV  
 HW  
 HX  
 HY  
 HZ  
 IA  
 IB  
 IC  
 ID  
 IE  
 IF  
 IG  
 IH  
 II  
 IJ  
 IK  
 IL  
 IM  
 IN  
 IO  
 IP  
 IQ  
 IR  
 IS  
 IT  
 IU  
 IV  
 IW  
 IX  
 IY  
 IZ  
 JA  
 JB  
 JC  
 JD  
 JE  
 JF  
 JG  
 JH  
 JI  
 JJ  
 JK  
 JL  
 JM  
 JN  
 JO  
 JP  
 JQ  
 JR  
 JS  
 JT  
 JU  
 JV  
 JW  
 JX  
 JY  
 JZ  
 KA  
 KB  
 KC  
 KD  
 KE  
 KF  
 KG  
 KH  
 KI  
 KJ  
 KK  
 KL  
 KM  
 KN  
 KO  
 KP  
 KQ  
 KR  
 KS  
 KT  
 KU  
 KV  
 KW  
 KX  
 KY  
 KZ  
 LA  
 LB  
 LC  
 LD  
 LE  
 LF  
 LG  
 LH  
 LI  
 LJ  
 LK  
 LL  
 LM  
 LN  
 LO  
 LP  
 LQ  
 LR  
 LS  
 LT  
 LU  
 LV  
 LW  
 LX  
 LY  
 LZ  
 MA  
 MB  
 MC  
 MD  
 ME  
 MF  
 MG  
 MH  
 MI  
 MJ  
 MK  
 ML  
 MM  
 MN  
 MO  
 MP  
 MQ  
 MR  
 MS  
 MT  
 MU  
 MV  
 MW  
 MX  
 MY  
 MZ  
 NA  
 NB  
 NC  
 ND  
 NE  
 NF  
 NG  
 NH  
 NI  
 NJ  
 NK  
 NL  
 NM  
 NN  
 NO  
 NP  
 NQ  
 NR  
 NS  
 NT  
 NU  
 NV  
 NW  
 NX  
 NY  
 NZ  
 OA  
 OB  
 OC  
 OD  
 OE  
 OF  
 OG  
 OH  
 OI  
 OJ  
 OK  
 OL  
 OM  
 ON  
 OO  
 OP  
 OQ  
 OR  
 OS  
 OT  
 OU  
 OV  
 OW  
 OX  
 OY  
 OZ  
 PA  
 PB  
 PC  
 PD  
 PE  
 PF  
 PG  
 PH  
 PI  
 PJ  
 PK  
 PL  
 PM  
 PN  
 PO  
 PP  
 PQ  
 PR  
 PS  
 PT  
 PU  
 PV  
 PW  
 PX  
 PY  
 PZ  
 QA  
 QB  
 QC  
 QD  
 QE  
 QF  
 QG  
 QH  
 QI  
 QJ  
 QK  
 QL  
 QM  
 QN  
 QO  
 QP  
 QQ  
 QR  
 QS  
 QT  
 QU  
 QV  
 QW  
 QX  
 QY  
 QZ  
 RA  
 RB  
 RC  
 RD  
 RE  
 RF  
 RG  
 RH  
 RI  
 RJ  
 RK  
 RL  
 RM  
 RN  
 RO  
 RP  
 RQ  
 RR  
 RS  
 RT  
 RU  
 RV  
 RW  
 RX  
 RY  
 RZ  
 SA  
 SB  
 SC  
 SD  
 SE  
 SF  
 SG  
 SH  
 SI  
 SJ  
 SK  
 SL  
 SM  
 SN  
 SO  
 SP  
 SQ  
 SR  
 SS  
 ST  
 SU  
 SV  
 SW  
 SX  
 SY  
 SZ  
 TA  
 TB  
 TC  
 TD  
 TE  
 TF  
 TG  
 TH  
 TI  
 TJ  
 TK  
 TL  
 TM  
 TN  
 TO  
 TP  
 TQ  
 TR  
 TS  
 TT  
 TU  
 TV  
 TW  
 TX  
 TY  
 TZ  
 UA  
 UB  
 UC  
 UD  
 UE  
 UF  
 UG  
 UH  
 UI  
 UJ  
 UK  
 UL  
 UM  
 UN  
 UO  
 UP  
 UQ  
 UR  
 US  
 UT  
 UU  
 UV  
 UW  
 UX  
 UY  
 UZ  
 VA  
 VB  
 VC  
 VD  
 VE  
 VF  
 VG  
 VH  
 VI  
 VJ  
 VK  
 VL  
 VM  
 VN  
 VO  
 VP  
 VQ  
 VR  
 VS  
 VT  
 VU  
 VV  
 VW  
 VX  
 VY  
 VZ  
 WA  
 WB  
 WC  
 WD  
 WE  
 WF  
 WG  
 WH  
 WI  
 WJ  
 WK  
 WL  
 WM  
 WN  
 WO  
 WP  
 WQ  
 WR  
 WS  
 WT  
 WU  
 WV  
 WW  
 WX  
 WY  
 WZ  
 XA  
 XB  
 XC  
 XD  
 XE  
 XF  
 XG  
 XH  
 XI  
 XJ  
 XK  
 XL  
 XM  
 XN  
 XO  
 XP  
 XQ  
 XR  
 XS  
 XT  
 XU  
 XV  
 XW  
 XX  
 XY  
 XZ  
 YA  
 YB  
 YC  
 YD  
 YE  
 YF  
 YG  
 YH  
 YI  
 YJ  
 YK  
 YL  
 YM  
 YN  
 YO  
 YP  
 YQ  
 YR  
 YS  
 YT  
 YU  
 YV  
 YW  
 YX  
 YY  
 YZ  
 ZA  
 ZB  
 ZC  
 ZD  
 ZE  
 ZF  
 ZG  
 ZH  
 ZI  
 ZJ  
 ZK  
 ZL  
 ZM  
 ZN  
 ZO  
 ZP  
 ZQ  
 ZR  
 ZS  
 ZT  
 ZU  
 ZV  
 ZW  
 ZX  
 ZY  
 ZZ  
 aa  
 ab  
 ac  
 ad  
 ae  
 af  
 ag  
 ah  
 ai  
 aj  
 ak  
 al  
 am  
 an  
 ao  
 ap  
 aq  
 ar  
 as  
 at  
 au  
 av  
 aw  
 ax  
 ay  
 az  
 ba  
 bb  
 bc  
 bd  
 be  
 bf  
 bg  
 bh  
 bi  
 bj  
 bk  
 bl  
 bm  
 bn  
 bo  
 bp  
 bq  
 br  
 bs  
 bt  
 bu  
 bv  
 bw  
 bx  
 by  
 bz  
 ca  
 cb  
 cc  
 cd  
 ce  
 cf  
 cg  
 ch  
 ci  
 cj  
 ck  
 cl  
 cm  
 cn  
 co  
 cp  
 cq  
 cr  
 cs  
 ct  
 cu  
 cv  
 cw  
 cx  
 cy  
 cz  
 da  
 db  
 dc  
 dd  
 de  
 df  
 dg  
 dh  
 di  
 dj  
 dk  
 dl  
 dm  
 dn  
 do  
 dp  
 dq  
 dr  
 ds  
 dt  
 du  
 dv  
 dw  
 dx  
 dy  
 dz  
 ea  
 eb  
 ec  
 ed  
 ee

Query Match 99.1%; Score 3600.5; DB 4; Length 665;  
 Best Local Similarity 99.2%; Pred. No. 1.4e-214;  
 Matches 652; Conservative 1; Mismatches 3; Indels 1; Gaps 1;

QY 1 MASTDSTYSQAQAQOQYSAVTAQPTQGYAQT-QAYGOOSYGTGYGPTDVSYTAQTTA 59  
 DB 9 MASTDSTYSQAQAQOQYSAVTAQPTQGYAQTQQAAGQOOSYGTGYGPTDVSYTAQTTA 68  
 QY 60 TYGQTAATSYGQPPPTGYTTPTAPQAYSQVQYGTGATYTTATTTTTOASTAAGSAYG 119  
 DB 69 TYGQTAATSYGQPPPTGYTTPTAPQAYSQVQYGTGATYTTATTTTTOASTAAGSAYG 128  
 QY 120 TOBAYPAGGQAPATAPTRPDGKPTETSDOPSTGYNOPSLGYSQSYSPQVGSY 179  
 DB 129 TOBAYPAGGQAPATCTRTEDANKPTETSDOPSTGYNOPSLGYSQSYSPQVGSY 188  
 QY 180 PMQPTAPSPPTSYSTPTSTYDOSYSQONTYQGPSSYGOOSYSGOOSYSGOQPTTS 239  
 DB 189 PMQPTAPSPPTSYSTPTSTYDOSYSQONTYQGPSSYGOOSYSGOOSYSGOQPTTS 248  
 QY 240 YPQTGSYQAPSPQYSQOOSYSGOOSYFRODHPSSMGVYQESGSGFSGPENTSMGPDN 299  
 DB 249 YPQTGSYQAPSPQYSQOOSYSGOOSYFRODHPSSMGVYQESGSGFSGPENTSMGPDN 308  
 QY 300 RGRGRGPDGRGSGRGGGSGRGGSGRGGSGRGGSGRGGSGRGGSGRGGSGRGGSGRGG 359  
 DB 309 RGRGRGPDGRGSGRGGGSGRGGSGRGGSGRGGSGRGGSGRGGSGRGGSGRGGSGRGG 368  
 QY 360 NSAIYVQGLNDSTYLDLADPFKQCGVVKNNKRTGQPMIHIYLDKETGPKGDAVSYED 419  
 DB 369 NSAIYVQGLNDSTYLDLADPFKQCGVVKNNKRTGQPMIHIYLDKETGPKGDAVSYED 428  
 QY 420 PPTAKAAVWFQDGFQSGSLKATSLARKKPPMNSMRGGLPRRGRGMPPLRGSGPGPG 479  
 DB 429 PPTAKAAVWFQDGFQSGSLKATSLARKKPPMNSMRGGLPRRGRGMPPLRGSGPGPG 488  
 QY 480 PGGPMRGMGRGGDRGGFPPRGRGSGRGNPSGGGNVQHRAGDWCPCPPGCGNPFARTE 539  
 DB 489 PGGPMRGMGRGGDRGGFPPRGRGSGRGNPSGGGNVQHRAGDWCPCPPGCGNPFARTE 548  
 QY 540 CNOCKAPKEGFLPPPPPPGDRGGRGPGGMRGGRGLMDRGPGGMPFRGGRGDRGF 599  
 DB 549 CNOCKAPKEGFLPPPPPPGDRGGRGPGGMRGGRGLMDRGPGGMPFRGGRGDRGF 608  
 QY 600 RGRGRGDRGGFQSGRGGPGGPPGLMEQWGRGGRGGKMDKGBHRQERDRPY 656  
 DB 609 RGRGRGDRGGFQSGRGGPGGPPGLMEQWGRGGRGGKMDKGBHRQERDRPY 665

RESULT 6  
 ABB57126  
 ID ABB57126 standard; protein; 655 AA.  
 AC ABB57126;  
 XX 07-MAR-2002 (first entry)  
 DE Mouse ischaemic condition related protein sequence SEQ ID NO:289.  
 XX  
 XX Mouse; ischaemia; compressive ischaemia; occlusive ischaemia;  
 KM vasoospastic ischaemia; ischaemic condition; ischaemic disease.  
 XX  
 OS Mus musculus.  
 XX  
 PN WO200188188-A2.  
 XX  
 PD 22-NOV-2001.  
 XX  
 PF 18-MAY-2001; 2001MO-JP04192.  
 XX  
 PR 18-MAY-2000; 2000JP-00145977.  
 XX  
 PA (UN1-) UNIT NIIHON SCHOOL JURIDICAL PERSON.

XX  
 PI Ishikawa K, Arai S, Takahashi Y, Nagata T, Ishii Y;  
 XX  
 DR WPI; 2002-034733/04.  
 DR N-PSDB; AB199383.  
 XX  
 PT Examining the ischemic condition (e.g. occlusive ischemia) by measuring  
 PT expression levels of particular genes defined in the specification or by  
 PT determining the expression profile of a gene group comprising these  
 PT genes.  
 PS Claim 2; Page 799-802; 2690pp; English.  
 XX  
 CC The present invention describes a method for examining ischaemic  
 CC conditions, comprising measuring the expression levels of particular  
 CC genes (i) in a test sample or determining the expression profile of a  
 CC gene group in the sample comprising genes selected from (i). The method  
 CC is useful for examining the ischaemic condition (e.g. compressive  
 CC ischaemia, occlusive ischaemia or vasoospastic ischaemia) by measuring  
 CC expression levels of particular genes (AB199202 to AB199912, encoding the  
 CC protein sequences in ABB57020 to ABB57374) or by determining the  
 CC expression profile of a gene group comprising these genes. The expression  
 CC levels or expression profiles produced by these genes are used as an  
 CC indicator when screening for ischaemic condition-improving drugs or  
 CC therapeutics for ischaemic diseases. AB199913 and AB199914 represent PCR  
 CC primers for a mouse ischaemic condition related sequence, which are used  
 CC in the exemplification of the present invention  
 CC  
 SQ Sequence 655 AA;

Query Match 98.3%; Score 3571.5; DB 5; Length 655;  
 Best Local Similarity 98.0%; Pred. No. 8.7e-213;  
 Matches 643; Conservative 8; Mismatches 4; Indels 1; Gaps 1;

QY 1 MASTDSTYSQAQAQOQYSAVTAQPTQGYAQTQQAAGQOOSYGTGYGPTDVSYTAQTTA 60  
 DB 1 MASTDSTYSQAQAQOQYSAVTAQPTQGYAQTQQAAGQOOSYGTGYGPTDVSYTAQTTA 60  
 QY 61 YGQTAATSYGQPPPTGYTTPTAPQAYSQVQYGTGATYTTATTTTTOASTAAGSAYG 120  
 DB 61 YGQTAATSYGQPPPTGYTTPTAPQAYSQVQYGTGATYTTATTTTTOASTAAGSAYG 120  
 QY 121 QPAPYVYGOQAPATATPRTDQGNKPTETSDOPSTGYNOPSLGYSQSYSPQVGSY 180  
 DB 121 QPAPYVYGOQAPATATPRTDQGNKPTETSDOPSTGYNOPSLGYSQSYSPQVGSY 180  
 QY 181 MOPVTAPSPPTSYSTPTSTYDOSYSQONTYQGPSSYGOOSYSGOOSYSGOQPTTS 240  
 DB 181 MOPVTAPSPPTSYSTPTSTYDOSYSQONTYQGPSSYGOOSYSGOOSYSGOQPTTS 240  
 QY 241 PPTGSYQAPSPQYSQOOSYSGOOSYFRODHPSSMGVYQESGSGFSGPENTSMGPDN 300  
 DB 241 PPTGSYQAPSPQYSQOOSYSGOOSYFRODHPSSMGVYQESGSGFSGPENTSMGPDN 300  
 QY 301 GGRGGRFDRGMSRGRGGGSGRGGSGRGGSGRGGSGRGGSGRGGSGRGGSGRGGSGRGG 359  
 DB 301 GGRGGRFDRGMSRGRGGGSGRGGSGRGGSGRGGSGRGGSGRGGSGRGGSGRGGSGRGG 360  
 QY 361 SAITYVQGLNDSTYLDLADPFKQCGVVKNNKRTGQPMIHIYLDKETGPKGDAVSYED 420  
 DB 361 SAITYVQGLNDSTYLDLADPFKQCGVVKNNKRTGQPMIHIYLDKETGPKGDAVSYED 420  
 QY 421 PPTAKAAVWFQDGFQSGSLKATSLARKKPPMNSMRGGLPRRGRGMPPLRGSGPGPG 480  
 DB 421 PPTAKAAVWFQDGFQSGSLKATSLARKKPPMNSMRGGLPRRGRGMPPLRGSGPGPG 480  
 QY 481 GGPMMGRGGRGGDRGGFPPRGRGSGRGNPSGGGNVQHRAGDWCPCPPGCGNPFARTEC 539  
 DB 481 GGPMMGRGGRGGDRGGFPPRGRGSGRGNPSGGGNVQHRAGDWCPCPPGCGNPFARTEC 539  
 QY 541 NCKKAPKEGFLPPPPPPGDRGGRGPGGMRGGRGLMDRGPGGMPFRGGRGDRGF 600  
 DB 541 NCKKAPKEGFLPPPPPPGDRGGRGPGGMRGGRGLMDRGPGGMPFRGGRGDRGF 600

QY 601 GGRGMDRGGRGGRRGGPGGPPGLMEOMGRRGGRRGGPGKMDKGHRORRRDPY 656  
 DB 600 GGRGMDRGGRGGRRGGPGGPPGLMEOMGRRGGRRGGPGKMDKGHRORRRDPY 655

## RESULT 7

ID ADAS5145 standard; protein; 600 AA.

ADAS5145;

20-NOV-2003 (first entry)

Human protein, SEQ ID 2713.

Cytostatic; Anti-inflammatory; Osteopapthic; Neuroprotective; Nootropic;  
 Gene Therapy; human; secretory protein; membrane proteins; cancer;  
 inflammatory disease; osteoporosis; neurological disease.

Homo sapiens.

EP1293569-A2.

19-MAR-2003.

21-MAR-2002; 2002EP-00006586.

14-SEP-2001; 2001JP-00328381.

24-JAN-2002; 2002US-0350435P.

(HELI-) HELIX RES INST.

(REAS-) RES ASSOC BIOTECHNOLOGY.

Isogai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S,  
 Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I,  
 Seki N, Yoshikawa T, Otsuka M, Nagahara K, Masuno Y;

WPI; 2003-395539/38.

N-PSDB; ADAS3506.

New polynucleotides encoding full-length polypeptides, e.g. secretory  
 and/or membrane proteins, useful for developing medicines for diseases in  
 which the gene is involved, or as target molecules for gene therapy.

Claim 14; SEQ ID NO 2713; 205pp; English.

The present invention relates to novel human secretory or membrane  
 proteins (ADAS4072-ADAS5710) and their coding sequences (ADAS2433-  
 CC ADAS4071). The coding sequences are useful in the gene therapy of  
 diseases caused by abnormalities of the proteins, e.g. cancer,  
 inflammatory diseases, osteoporosis or neurological disease.

Sequence 600 AA;

Query Match 90.2%; Score 3277; DB 6; Length 600;  
 Best Local Similarity 91.3%; Pred. No. 1.3e-194;  
 Matches 599; Conservative 0; Mismatches 1; Indels 56; Gaps 1;

QY 1 MASTDSTYSQAQAAGQYSAYTAQPTGVAQTQAYGQSGYGTGPTDVSYTAQTAT 60  
 DB 1 MASTDSTYSQAQAAGQYSAYTAQPTGVAQTQAYGQSGYGTGPTDVSYTAQTAT 60  
 QY 61 YGCTAATATSGGPTGTTTAPQAYSQPIQYGTGAYDTTATVTTTQASVAAAGAYGT 120  
 DB 61 YGCTAATATSGGPTGTTTAPQAYSQPIQYGTGAYDTTATVTTTQASVAAAGAYGT 120  
 QY 121 QPAYPYGQOPATAPTRPDGKPKRTSQPGSTGYNQPSLGYGQSNVSYQVPGSY 180  
 DB 121 QPAYPYGQOPATAPTRPDGKPKRTSQPGSTGYNQPSLGYGQSNVSYQVPGSY 180  
 QY 181 MQPVTAPPSYPTSYSTQPTSYDQSSYQONTYGGQSSYGGQSSYGGQSSYGGQSSY 240  
 DB 181 MQPVTAPPSYPTSYSTQPTSYDQSSYQONTYGGQSSYGGQSSYGGQSSYGGQSSY 240

DB 136 -----PTSYSTQPTSYDQSSYQONTYGGQSSYGGQSSYGGQSSYGGQSSY 184  
 QY 241 PPTGTSYQAPSOYQOQSSYQOQSSYQOQSSYQOQSSYQOQSSYQOQSSYQOQSSY 300  
 DB 185 PPTGTSYQAPSOYQOQSSYQOQSSYQOQSSYQOQSSYQOQSSYQOQSSYQOQSSY 244  
 QY 301 GRRGGRFDRGMSRRGRRGGRRGGRRGGRRGGRRGGRRGGRRGGRRGGRRGGRRGG 360  
 DB 245 GRRGGRFDRGMSRRGRRGGRRGGRRGGRRGGRRGGRRGGRRGGRRGGRRGGRRGG 304  
 QY 361 SAIYVQGLNDSYTLDDLADFPKQCGVYKNNKRTGQPMIHITLDKTKGKRGATVSE 420  
 DB 305 SAIYVQGLNDSYTLDDLADFPKQCGVYKNNKRTGQPMIHITLDKTKGKRGATVSE 364  
 QY 421 PTKAAVEMFDDKDRPGSKLKVSLAKKPPNNMSMGGLPPRRGGMPPLRGPGGPGG 480  
 DB 365 PTKAAVEMFDDKDRPGSKLKVSLAKKPPNNMSMGGLPPRRGGMPPLRGPGGPGG 424  
 QY 481 GSPMGRMGRRGGDRGFPFRGRRGRRGGRRGGRRGGRRGGRRGGRRGGRRGGRRGG 540  
 DB 425 GSPMGRMGRRGGDRGFPFRGRRGRRGGRRGGRRGGRRGGRRGGRRGGRRGGRRGG 484  
 QY 541 NOCKAPKPEGLPPPPPPGSDRGRRGGRRGGRRGGRRGGRRGGRRGGRRGGRRGG 600  
 DB 485 NOCKAPKPEGLPPPPPPGSDRGRRGGRRGGRRGGRRGGRRGGRRGGRRGGRRGG 544  
 QY 601 GGRGMDRGGRGGRRGGPGGPPGLMEOMGRRGGRRGGPGKMDKGHRORRRDPY 656  
 DB 545 GGRGMDRGGRGGRRGGPGGPPGLMEOMGRRGGRRGGPGKMDKGHRORRRDPY 600

## RESULT 8

ID ADAS4072 standard; protein; 604 AA.

ADAS4072;

04-NOV-2004 (first entry)

Human protein useful for treating neurological disease Seq 2908.

human; oligo-capping method; diagnostic marker; gene therapy;  
 osteoporosis; neurological disease; Alzheimer's disease;  
 Parkinson's disease; dementia; short memory; cancer;

gene or motor function; emotional reaction; fear response; panic;  
 osteopapthic; neuroprotective; nootropic; antiparkinsonian; cyostatic;  
 tranquilizer.

Homo sapiens.

EP1447413-A2.

18-AUG-2004.

12-FEB-2004; 2004EP-00003145.

14-FEB-2003; 2003JP-00102207.

09-MAY-2003; 2003JP-00131452.

(REAS-) RES ASSOC BIOTECHNOLOGY.

Isogai T, Yamamoto J, Nishikawa T, Isono Y, Sugiyama T, Otsuki T,  
 Makamatsu A, Ishii S, Nagai K, Irie R;

WPI; 2004-583265/57.

N-PSDB; ADAS7446.

New 1995 cDNA, useful for treating osteoporosis, neurological diseases,  
 Alzheimer's diseases, Parkinson's diseases, dementia and various cancers.

Claim 1; SEQ ID NO 2908; 266pp; English.

This invention relates to novel, isolated full length human cDNA

CC

molecules and the encoded proteins thereof. Specifically, it refers to CDNA clones obtained by an oligo-capping method, where none of these clones are identical to any known human mRNAs. The present invention describes an immunosassay to identify agonists and antagonists, as well as antibodies, antisense molecules and siRNAs that can all be used to bind to and modulate expression of the CDNA molecules. As such, these molecules are useful for diagnostic markers or therapeutic targets for the various diseases or morbid states. In particular, they are useful in gene therapy for treating osteoporosis, neurological disease, Alzheimer's disease, Parkinson's disease, dementia, short memory and various cancers, as well as for maintaining equilibrium of sense or motor function, and for treating emotional reaction, fear response and panic. Accordingly, they exhibit osteoprotective, neuroprotective, nootropic, antiparkinsonian, cytostatic and tranquilizer activities. This polypeptide is a protein encoded by a full length human CDNA sequence of the invention. NOTE: This sequence is not given in the sequence listing of the specification but can be obtained on CD-ROM from the European Patent Office, Vienna Sub-office.

Sequence 604 AA:

Query Match 88.4%; Score 3210; DB 8; Length 604;  
Best Local Similarity 89.9%; Pred. No. 1.9e-190;  
Matches 590; Conservative 3; Mismatches 11; Indels 52; Gaps 4;

1 MASTDSTYSQAQAAGQSYAATQPTQAGTQAGQSYGYGQPTDSTYQAQTAT 60  
1 MASTDSTYSQAQAAGQSYAATQPTQAGTQAGTQAGQSYGYGQPTDSTYQAQTAT 60  
61 YGQTATATSYGQPTGTTTPTAPQAYGQPVQYGTGAYDPTTATVTTTQSYAQAQSYGT 120  
61 YGQTATATSYGQPTGTTTPTAPQAYGQPVQYGTGAYDPTTATVTTTQSYAQAQSYGT 120  
121 QPAPYPAVYQQAATAPTRPDQGNKPTETQPSSTGTYNPSLQYGSNYSYQVPSY 180  
100 -----YSQOPAPATAPTRPDQGNKPTETQPSSTGTYNPSLQYGSNYSYQVPSY 153  
181 MOPVTPAPSYPTSYSTQPTSTYDQSSYQANTYGPQSSYQOQSSYQOQPTSY 240  
154 MOPVTPAPSYPT-----SYQOQSSYQOQSSYQOQPTSY 188  
241 PPTQTSYQAQPSYQOQSSYQOQSSYQOQSSYQOQSSYQOQSSYQOQSSYQOQSSY 300  
189 PPTQTSYQAQPSYQOQSSYQOQSSYQOQSSYQOQSSYQOQSSYQOQSSYQOQSSY 248  
301 GRRGGSFDRGMSRGRGGRGMSRGRGGRGMSRGRGGRGMSRGRGGRGMSRGRGGR 360  
249 GRRGGSFDRGMSRGRGGRGMSRGRGGRGMSRGRGGRGMSRGRGGRGMSRGRGGR 308  
361 SAIYVGLDSTVLDLADLFKCCGYVKNKRTGQPMIHTYLDKFGKPKGATVSYEDP 420  
309 SAIYVGLDSTVLDLADLFKCCGYVKNKRTGQPMIHTYLDKFGKPKGATVSYEDP 368  
421 PPAKAAVEMFDGDFGSKLKVSLARKKPPMNSMRGGLPRREGMPPLRGAGRGPG 480  
369 PPAKAAVEMFDGDFGSKLKVSLARKKPPMNSMRGGLPRREGMPPLRGAGRGPG 428  
481 GGPWGMGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGR 540  
429 GGPWGMGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGR 488  
541 NQCKAKPESGFLPPLPPPGGDDGRRGPGCMRGGRGLMDRGPGMFRGGRGGRGGR 600  
489 NQCKAKPESGFLPPLPPPGGDDGRRGPGCMRGGRGLMDRGPGMFRGGRGGRGGR 548  
601 GGRGMDRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGR 656  
549 GGRGMDRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGR 604

RESULT 9  
ADP56334  
ID ADP56334 standard; protein; 583 AA.

XX AC ADP56334;  
XX DT 18-NOV-2004 (first entry)  
XX DE Human PRO protein sequence SEQ ID NO:2310.  
XX KW human; PRO; immune related disease; inflammatory immune response;  
XX KW immune response stimulation; antiallergic; antianaemic; antiarthritic;  
XX KW antidiabetic; antidiabetic; antiinflammatory; antipsoriatic;  
XX KW antirheumatic; antithyroid; CNS; dermatological; gastrointestinal;  
XX KW haemostatic; hepatotropic; immunosuppressant; immunosuppressive; muscular;  
XX KW nephrotropic; neuroprotective; osteopathic; respiratory; vasotropic;  
XX KW virucide; gene therapy.  
XX OS Homo sapiens.  
XX PN NC02004039956-A2.  
XX PD 13-MAY-2004.  
XX PF 28-OCT-2003; 2003MO-US034381.  
XX PR 29-OCT-2002; 2002US-0422472P.  
XX PA (GENT ) GENENTECH INC.  
XX PI Agarwal S, Clark H, Gurney AL, Schoenfeld J, Williams PM;  
XX PI Wood WI, Wu TD;  
XX DR WPI, 2004-376182/35.  
XX DR N-PSDB; ADP56333.  
XX PT New PRO polynucleotides and polypeptides, useful in diagnosing  
XX PT and treating an immune related disease, e.g. systemic lupus  
XX PT erythematosus, rheumatoid arthritis, diabetes mellitus or asthma and in  
XX PT stimulating an immune response.  
XX PS Claim 1; SEQ ID NO 2310; 3009pp; English.  
XX CC The present invention describes an isolated PRO nucleic acid (I). Also  
XX CC described: (1) a vector comprising (1); (2) a host cell comprising the  
XX CC vector of (1); (3) a process for producing a PRO polypeptide; (4) an  
XX CC isolated PRO polypeptide; (5) a chimeric molecule comprising the  
XX CC polypeptide of (4) fused to a heterologous amino acid sequence; (6) an  
XX CC antibody which specifically binds to a polypeptide of (4); (7) a  
XX CC composition of matter comprising a polypeptide of (4), an agonist or  
XX CC antagonist of the polypeptide or an antibody that binds to the  
XX CC polypeptide in combination with a carrier; (8) an article of manufacture  
XX CC comprising a container, a label on the container and a composition of  
XX CC matter of (7); (9) a method of treating an immune related disease in a  
XX CC mammal; (10) a method for determining the presence of a PRO polypeptide  
XX CC in a sample suspected of having the polypeptide; (11) a method of  
XX CC diagnosing an immune related disease or an inflammatory immune response  
XX CC in a mammal; (12) a method of identifying a compound that inhibits or  
XX CC mimics the activity of or expression of a gene encoding a PRO polypeptide  
XX CC; and (13) a method of stimulating the immune response in a mammal. The  
XX CC PRO sequences have antiallergic, antianaemic, antiarthritic,  
XX CC antidiabetic, antidiabetic, antiinflammatory, antipsoriatic,  
XX CC antirheumatic, antithyroid, CNS, dermatological, gastrointestinal,  
XX CC haemostatic, hepatotropic, immunosuppressant, immunosuppressive, muscular,  
XX CC nephrotropic, neuroprotective, osteopathic, respiratory, vasotropic and  
XX CC virucide activities, and can be used in gene therapy. The nucleic acid  
XX CC (I) and the encoded polypeptides, compositions, kits and methods are  
XX CC useful in diagnosing and treating an immune related disease and in  
XX CC stimulating an immune response. The present sequence represents a human  
XX CC PRO protein from the present invention.  
XX SQ Sequence 583 AA;

Query Match 87.5%; Score 3178.5; DB 8; Length 583;  
Best Local Similarity 88.7%; Pred. No. 1.6e-188;  
Matches 582; Conservative 0; Mismatches 1; Indels 73; Gaps 1;





ID ABG95068 standard; protein; 476 AA.  
 XX  
 AC ABG95068;  
 DT 04-DEC-2002 (first entry)  
 XX  
 DE Human translocation (11; 22) (q24; q12) protein #1.  
 XX  
 KM Chromosome aberration; oncogenic fusion protein; cancer;  
 KM proliferative disease; cellular protein isoform; heat shock protein 90;  
 KM HSP-90; rheumatoid arthritis; cancer; hematopoietic disorder;  
 KM T cell lymphoma; B cell lymphoma; chronic myeloid leukaemia; CML;  
 KM acute myeloid leukaemia; AML; chronic myelomonocytic leukaemia; CMML;  
 KM acute lymphoblastic leukaemia; ALL; APL; NHL; solid tumour;  
 KM papillary thyroid carcinoma; Ewing's sarcoma; melanoma; liposarcoma;  
 KM rhabdomyosarcoma; synovial sarcoma; viral infection.  
 XX  
 OS Homo sapiens.  
 XX  
 PN MO200269900-A2.  
 XX  
 PD 12-SEP-2002.  
 XX  
 PF 01-MAR-2002; 2002MO-US006518.  
 XX  
 PR 01-MAR-2001; 2001US-0272751P.  
 XX  
 PA (CONF-) CONFORMA THERAPEUTICS CORP.  
 XX  
 PI Fritz LC, Burrows FJ;  
 XX  
 DR WPI, 2002-698710/75.  
 XX  
 DR N-PSDB; ABS73252.  
 XX  
 PT Treating genetically-defined disease associated with chromosomal  
 PT aberrations yielding oncogenic fusion proteins, e.g. cell proliferative  
 PT diseases, involves administering an inhibitor of heat shock protein 90.  
 XX  
 PS Disclosure; Page 204-206; 389pp; English.  
 XX  
 XX The invention describes a method of treating genetically-defined disease  
 CC associated with chromosomal aberrations yielding oncogenic fusion  
 CC proteins (I), treating cancerous cells containing (I) in a heterogeneous  
 CC cell population, treating proliferative diseases associated with mutant  
 CC protein or cellular protein isoforms (II) dependent on heat shock protein  
 CC (HSP)-90, or selectively treating cells expressing (II) involving  
 CC administering HSP90-inhibitor. The method is useful for treating  
 CC genetically-defined disease with chromosomal aberration yielding  
 CC oncogenic fusion protein, treating cancerous cells containing fusion  
 CC protein in heterogeneous cell population, treating proliferative disease  
 CC (e.g. rheumatoid arthritis or cancer) associated with mutant protein or  
 CC cellular protein isoform dependent on heat shock protein (HSP)-90 (e.g.  
 CC p53), or selectively treating cells expressing mutant protein or cellular  
 CC protein isoform in a patient heterozygous for (II). The method is useful  
 CC for treating a disease e.g. hematopoietic disorder such as T or B cell  
 CC lymphoma, chronic myeloid leukaemia (CML), APL, ALL, NHL and CMML,  
 CC or a disease characterised by a solid tumour such as papillary thyroid  
 CC carcinoma, Ewing's sarcoma, melanoma, liposarcoma, rhabdomyosarcoma and  
 CC synovial sarcoma. The method is also useful for treating viral  
 CC infections. This represents a protein encoded by the DNA sequence of a  
 CC chromosome aberration  
 CC  
 XX  
 SQ Sequence 476 AA;  
 XX  
 QY Query Match 39.3%; Score 1429; DB 5; Length 476;  
 DB Best Local Similarity 88.8%; Pred. No. 2,6e-80;  
 Mismatches 270; Conservative 6; Mismatches 16; Indels 12; Gaps 1;  
 QY 1 MASTDVTSTYSGAQAAGQSYSAVTAQPTGVAQTTOAVGQSGSYGTYGPTDVSYTQAQTAT 60  
 DB 1 MASTDVTSTYSGAQAAGQSYSAVTAQPTGVAQTTOAVGQSGSYGTYGPTDVSYTQAQTAT 60  
 QY 61 YGQTAATATSTGQPPPTGTTTPTAQAASQPVQSGTGTGAIVDTTATVTTTQASTAAGSAVGT 120

DB 61 YGQTAATATSTGQPPPTGTTTPTAQAASQPVQSGTGTGAIVDTTATVTTTQASTAAGSAVGT 120  
 QY 121 QPAVPAVGQPPAATATPPTPDGDKKPTETSGQPSSTGGYNQPSLGYGGSNYSYPOVPSSYP 180  
 DB 121 QPAVPAVGQPPAATATPPTPDGDKKPTETSGQPSSTGGYNQPSLGYGGSNYSYPOVPSSYP 180  
 QY 181 MCPVTAPPSYPPPSYSTQPTSTYDQSSYSQONTYGGPSSYGGQSSYGGQSPPTSY 240  
 DB 181 MCPVTAPPSYPPPSYSTQPTSTYDQSSYSQONTYGGPSSYGGQSSYGGQSPPTSY 240  
 QY 241 PPQTGSYSAQPSQYSGQSSSYGGQSS-----SFRQDHPSSWGVYGGSSGSP 288  
 DB 241 PPQTGSYSAQPSQYSGQSSSYGGQSSPPGGAOTITKXTEKRPQPDYQIIGPTSSRLANP 300  
 QY 289 GGNR 292  
 DB 301 GSGC 304  
 RESULT 12  
 AAM34972  
 ID AAM34972 standard; protein; 362 AA.  
 XX  
 AC AAM34972;  
 XX  
 DT 09-MAR-1998 (first entry)  
 XX  
 DE Chimeric Ewing's sarcoma-WT1 protein splice variant 2.  
 XX  
 KM Ewing's sarcoma; EMS; EMS-WT1 protein; peripheral neuroectodermal tumour;  
 KM ENET; breakpoint locus; Wilms' tumour;  
 KM desmoplastic small round cell tumour; DSRC tumour.  
 XX  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT Protein 1..264  
 FT /note="EMS protein"  
 FT 265..362  
 FT /note="WT1 protein"  
 XX  
 PN US5670317-A.  
 XX  
 PD 23-SEP-1997.  
 XX  
 PF 08-MAY-1995; 95US-00437027.  
 XX  
 PR 08-MAY-1995; 95US-00437027.  
 XX  
 PA (SLOK) SLOAN KETTERING INST CANCER RES.  
 XX  
 PI Ladanly M, Gerald W;  
 XX  
 DR WPI, 1997-479448/44.  
 XX  
 DR N-PSDB; AAT97870.  
 XX  
 PT diagnosis of desmoplastic small round cell tumours - by detecting nucleic  
 PT acid encoding chimeric EMS-WT1 protein.  
 XX  
 PS Disclosure; Col 39-42; 34pp; English.  
 XX  
 XX The present sequence represents a chimeric human Ewing's sarcoma (EMS) -  
 CC WT1 protein. EMS is also known as peripheral neuroectodermal tumour  
 CC (ENET). The protein is composed of EMS and WT1 proteins (see features  
 CC table). The WT1 gene encodes 4 alternatively spliced RNA transcripts that  
 CC produce functionally different products. One alternative splice site is  
 CC at the 3' end of exon 9 and leads to products that differ by 3 amino  
 CC acids placed between the third and fourth zinc fingers of the DNA binding  
 CC domain. The present protein does not contain these three amino acids  
 CC (Lys, Thr, Ser), while AAM34971 does. Detection of EMS-WT1 can be  
 CC accomplished by reverse transcriptase PCR, where WT1 is screened (see  
 CC AAT97852-68) as a breakpoint locus because of its involvement in Wilms'

CC tumour, which shares some histopathologic features of desmoplastic small  
CC round cell (DSRC) tumours. The EMS-WT1 chimeric transcript has been  
CC detected in 11 out of 12 DSRC tumours studied and in none of 49 other  
CC tumours. DSRC tumours are associated with translocation of the EMS gene.  
CC Oligonucleotides based on the AAT97869 are used in a method for the  
CC diagnosis of DSRC tumours in patients. The method comprises detecting a  
CC nucleic acid molecule encoding a chimeric EMS-WT1 protein in a sample  
CC from the subject, where positive detection indicates the presence of a  
CC DSRC tumour

XX Sequence 362 AA;

Query Match 39.1%; Score 1420; DB 2; Length 362;

Best Local Similarity 100.0%; Pred. No. 6.9e-80; Mismatches 0; Gaps 0;

Matches 265; Conservative 0; Indels 0; Gaps 0;

QY 1 MASTDYSTYSGAQAAGGYSAYTAQPTGVAQTTOAYGQSSYGYGPTDVSYTOAQTAT 60  
DB 1 MASTDYSTYSGAQAAGGYSAYTAQPTGVAQTTOAYGQSSYGYGPTDVSYTOAQTAT 60  
QY 61 YGOTAVATSYGQPTGTTTPAPQAYSQPVQGYGTGAVIDTTTATVTTTQASVAAQSAVGT 120  
DB 61 YGOTAVATSYGQPTGTTTPAPQAYSQPVQGYGTGAVIDTTTATVTTTQASVAAQSAVGT 120  
QY 121 QPAYPAVGGQPPAATAPTRPDGKPTETSPQSSSTGYNQPSLGYGQSNYSYPOVPSY 180  
DB 121 QPAYPAVGGQPPAATAPTRPDGKPTETSPQSSSTGYNQPSLGYGQSNYSYPOVPSY 180  
QY 181 MQPVTAAPSYPPTSYSTQPTSYDQSSYQONTYGGPSSYGGQSSYGGQSSYGGQPTPTSY 240  
DB 181 MQPVTAAPSYPPTSYSTQPTSYDQSSYQONTYGGPSSYGGQSSYGGQSSYGGQPTPTSY 240  
QY 241 PRQTSYSQAPSPQSSYGGQSSYGGQSSYGGQSSYGGQSSYGGQSSYGGQSSYGGQSSY 265  
DB 241 PRQTSYSQAPSPQSSYGGQSSYGGQSSYGGQSSYGGQSSYGGQSSYGGQSSYGGQSSY 265

#### RESULT 13

AAW34971 standard; protein; 365 AA.

AC AAW34971;

DT 09-MAR-1998 (first entry)

XX Chimeric Ewing's sarcoma-WT1 protein splice variant 1.

XX Ewing's sarcoma; EMS; EMS-WT1 protein; peripheral neuroectodermal tumour;

KW PNT; breakpoint locus; Wilms' tumour;

KW desmoplastic small round cell tumour; DSRC tumour.

OS Homo sapiens.

XX Key Location/Qualifiers

FT Protein 1..264

FT Protein /note= "EMS protein"

FT Protein 265..365

FT Protein /note= "WT1 protein"

XX US5670317-A.

XX 23-SEP-1997.

XX 08-MAY-1995; 95US-00437027.

XX 08-MAY-1995; 95US-00437027.

XX (SLOK) SLOAN KETTERING INST CANCER RES.

XX Ledanyi M, Gerald W;

XX WPI; 1997-479448/44.

XX N-PSDB; AAT97869.

XX diagnosis of desmoplastic small round cell tumours - by detecting nucleic  
PT acid encoding chimeric EMS-WT1 protein.

XX Disclosure; Col 37-40; 34pp; English.

XX The present sequence represents a chimeric human Ewing's sarcoma (EMS) -  
CC WT1 protein. EMS is also known as peripheral neuroectodermal tumor  
CC (PNET). The protein is composed of EMS and WT1 proteins (see features  
CC table). The WT1 gene encodes 4 alternatively spliced RNA transcripts that  
CC produce functionally different products. One alternative splice site is  
CC at the 3' end of exon 9 and leads to products that differ by 3 amino  
CC acids placed between the third and fourth zinc fingers of the DNA binding  
CC domain. The present protein contains these three amino acids (Lys, Thr,  
CC Ser), while AAW34972 does not. Detection of EMS-WT1 can be accomplished  
CC by reverse transcriptase PCR, where WT1 is screened (see AAT97852-68) as  
CC a breakpoint locus because of its involvement in Wilms' tumour, which  
CC shares some histopathologic features of desmoplastic small round cell  
CC (DSRC) tumours. The EMS-WT1 chimeric transcript has been detected in 11  
CC out of 12 DSRC tumours studied and in none of 49 other tumours. DSRC  
CC tumours are associated with translocation of the EMS gene.  
CC Oligonucleotides based on the AAT97869 are used in a method for the  
CC diagnosis of DSRC tumours in patients. The method comprises detecting a  
CC nucleic acid molecule encoding a chimeric EMS-WT1 protein in a sample  
CC from the subject, where positive detection indicates the presence of a  
CC DSRC tumour

XX Sequence 365 AA;

Query Match 39.1%; Score 1420; DB 2; Length 365;

Best Local Similarity 100.0%; Pred. No. 6.9e-80; Mismatches 0; Gaps 0;

Matches 265; Conservative 0; Indels 0; Gaps 0;

QY 1 MASTDYSTYSGAQAAGGYSAYTAQPTGVAQTTOAYGQSSYGYGPTDVSYTOAQTAT 60  
DB 1 MASTDYSTYSGAQAAGGYSAYTAQPTGVAQTTOAYGQSSYGYGPTDVSYTOAQTAT 60  
QY 61 YGOTAVATSYGQPTGTTTPAPQAYSQPVQGYGTGAVIDTTTATVTTTQASVAAQSAVGT 120  
DB 61 YGOTAVATSYGQPTGTTTPAPQAYSQPVQGYGTGAVIDTTTATVTTTQASVAAQSAVGT 120  
QY 121 QPAYPAVGGQPPAATAPTRPDGKPTETSPQSSSTGYNQPSLGYGQSNYSYPOVPSY 180  
DB 121 QPAYPAVGGQPPAATAPTRPDGKPTETSPQSSSTGYNQPSLGYGQSNYSYPOVPSY 180  
QY 181 MQPVTAAPSYPPTSYSTQPTSYDQSSYQONTYGGPSSYGGQSSYGGQSSYGGQPTPTSY 240  
DB 181 MQPVTAAPSYPPTSYSTQPTSYDQSSYQONTYGGPSSYGGQSSYGGQSSYGGQPTPTSY 240  
QY 241 PRQTSYSQAPSPQSSYGGQSSYGGQSSYGGQSSYGGQSSYGGQSSYGGQSSYGGQSSY 265  
DB 241 PRQTSYSQAPSPQSSYGGQSSYGGQSSYGGQSSYGGQSSYGGQSSYGGQSSYGGQSSY 265

#### RESULT 14

AD126113 standard; protein; 518 AA.

AC AD126113;

DT 22-APR-2004 (first entry)

XX Human protein that promotes STAT6 activation #39.

XX human; signal transducer and activator of transcription 6; STAT6;

KW immunogen; STAT6 activation; allergy; inflammation; autoimmune disease;

KW diabetes; hyperlipidaemia; infection; cancer; Th1 hyperactive disease;

KW rheumatoid arthritis; osteoarthritis; systemic lupus erythematosus;

KW sepsis; asthma; allergic rhinitis; ischaemic heart disease;

KW subarachnoid haemorrhage; viral hepatitis; AIDS.

XX Homo sapiens.

PN WO2003104277-A2.  
 XX  
 PD 18-DEC-2003.  
 XX  
 PF 05-JUN-2003; 2003MO-JP007123.  
 XX  
 PR 05-JUN-2002; 2002JP-00164257.  
 PR 06-JUN-2002; 2002JUS-0385912P.  
 PR 26-DEC-2002; 2002JP-00377326.  
 PR 27-DEC-2002; 2002JUS-0436467P.  
 PR 15-MAY-2003; 2003JP-00137505.  
 PR 16-MAY-2003; 2003JUS-0470836P.  
 XX  
 PA (ASAH) ASAH KASEI KK.  
 PI Sugahara T, Matsuda A, Honda G, Muramatsu S, Ishizawa K;  
 DR N-PSDB; AD126112.  
 XX  
 PT New signal transducer and activator of transcription 6 activation  
 PT promoting purified protein, for diagnosing and treating disease  
 PT associated with activation/inhibition of transcription factor e.g.  
 PT diabetes and cancer.  
 XX  
 PS Claim 1, SEQ ID NO 78; 1368bp; English.  
 XX  
 CC The invention relates to a purified protein promoting signal transducer  
 CC and activator of transcription 6 activation (STAT6). The protein is  
 CC useful for the producing an antibody, which involves administering the  
 CC protein or its epitope-bearing fragments to a non-human animal as an  
 CC antigen. The nucleic acid is useful for diagnosing a disease or  
 CC susceptibility to a disease related to expression or activity of the  
 CC protein. A transformant expressing the protein is useful for screening  
 CC compounds which inhibit or promote STAT6 activation. A transformant  
 CC expressing the protein is useful for producing a pharmaceutical  
 CC composition. Compositions, antibodies and antisense molecules are useful  
 CC for the treating a disease associated with STAT6 activation such as  
 CC allergic diseases, inflammation, autoimmune diseases, diabetes,  
 CC hyperlipidaemia, infectious disease and cancers. Compositions are useful  
 CC for treating disease associated with STAT6 activation and/or prevention  
 CC of Th1 hyperactive diseases. Compositions are also useful in rheumatoid  
 CC arthritis, osteoarthritis, systemic lupus erythematosus, sepsis, asthma,  
 CC allergic rhinitis, ischaemic heart diseases, subarachnoid haemorrhage,  
 CC viral hepatitis and AIDS. The protein has efficient promoting STAT6  
 CC activity. The protein or nucleic acid is effectively useful for screening  
 CC compounds for treating and preventing disease associated with excessive  
 CC activation or inhibition of STAT6. The present sequence represents the  
 CC amino acid sequence of a human protein which promotes STAT6 activation.  
 XX  
 SQ Sequence 518 AA;  
 Query Match 35.5%; Score 1290; DB 8; Length 518;  
 Best Local Similarity 43.9%; Pred. No. 1,1e-71;  
 Matches 301; Conservative 59; Mismatches 128; Indels 198; Gaps 30;  
 QY 1 MASTDYSYSAQAAGQGYSAHTAQTGVA-QTTQAVGQGSYGTGPGTVDVSYTAQTGA 59  
 DB 1 MASNDDY-----QATGSGVAYPTQPGGYSQSSQPGGQSYSGSAD-----TS 48  
 QY 60 TYGQTAYATSYGQPP-TGYTTTPAIPAAYISQPVQYG-TGAYDITTAATVTTTQASIAAQA 117  
 DB 49 GYQSSSYGSSYGQTQNTGYGTQSNP-----QGYGTGTYG-----SSQSSQSS 91  
 QY 118 YGTQPAIPAYAGQPPAATAPRPPDGNKFTENSOPTSGTGNGPSLGGSNYSYPPVPG 177  
 DB 92 YGQSSSYFGYGQPPA-----PSSYSGSYG-----G 116  
 QY 178 SYPMQVPTAPPSYPTSYSTOPTSYDQSSYQNTYGPSS--YGQSSY-GQSSSYGQ 234  
 DB 117 S-----SQSSSYGQPGQSGGYGQSGYGGQSSYGG 146  
 QY 235 QPPTSYPTQGSYQABSYQSSSYGQSSFRQDHPSPKMGVYGGSGFSGPGRKSM 294

Db 147 QSSSYNPQ--GYGQ-QNQNSSSGGGGGGG-----GNYGQDQSSMSGGGGGG 194  
 QY 295 SGPDNRGRGRGF-----DRGMSRG-----GGGGRGKGAG--ERGG 332  
 Db 195 GNQDQSGGGGGGGYGGQDQGRGGRGGGGGYNSSGGEYBRGRGGGGGGGGSDRG 254  
 QY 333 FNRKGPMDGEPPLDGPVDPEDSDNSAIYVOGLNDSTLDDLADFQCCVYKNNK 392  
 Db 255 FNRKGPMDGESSHD-----SEQDSDNNNTFVOGIGENVTIISVADYFQIITKNNK 309  
 QY 393 TQPMTHIYDKETKPKGDAATVSEDPPTAKAAVWFDKDPQGSKLARKKPPMN 452  
 Db 310 TQPMTHIYDKETKPKGDAATVSEDPPTAKAAVWFDKDPQGSKLARKKPPMN 369  
 QY 453 SMRGLPPEBRGMPPLLRGPGGPGGPGGPGGPGGPGGPGGPGGPGGPGGPGG 512  
 Db 370 --RGGNGRGRG-----RGGPGGPGGPGGPGGPGGPGGPGGPGGPGGPGG 409  
 QY 513 GNYQRAAGMOCNPGCGNPFARTECNCKAPKTBGFLPPPPPPGGRGGRGPGGMR 572  
 Db 410 GGGQQRAGMCKCPNPTCENNMFWRBECNQCKAPKPDG---PGGPGGSHMGANYGDDR 465  
 QY 573 GGRGLMDRGPGGMRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRG 631  
 Db 466 RGRGG-YDRGG---YR-GRGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGR 498  
 QY 632 RRGGRGPGGMRD-KGEHROERRDRPY 656  
 Db 499 -----GPGMRDGRGHRDRRRRPY 518  
 RESULT 15  
 AAM78355  
 ID AAM78355 standard; protein; 526 AA.  
 XX  
 AC AAM78355;  
 XX  
 DT 06-NOV-2001 (first entry)  
 XX  
 DE Human protein SEQ ID NO 1017.  
 XX  
 KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;  
 KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
 KW tissue growth factor; immunomodulatory; cancer; leukaemia;  
 KW nervous system disorder; arthritis; inflammation.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200157190-A2.  
 XX  
 PD 09-AUG-2001.  
 XX  
 PF 05-FEB-2001; 2001MO-US004098.  
 XX  
 PR 03-FEB-2000; 2000US-00496914.  
 PR 27-APR-2000; 2000US-00560875.  
 PR 20-JUN-2000; 2000US-00598075.  
 PR 19-JUL-2000; 2000US-00620325.  
 PR 01-SEP-2000; 2000US-00654936.  
 PR 15-SEP-2000; 2000US-00663561.  
 PR 20-OCT-2000; 2000US-00693325.  
 PR 30-NOV-2000; 2000US-00728422.  
 XX  
 PA (HYSE-) HYSEQ INC.  
 PI Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y;  
 PI Ma Y, Zhao QA, Wang D, Zhang J, Zhang J, Ren F, Chen R, Wang ZW;  
 PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;  
 XX  
 DR WPI; 2001-476283/51.  
 DR N-PSDB; AAK51488.

PT Nucleic acids encoding polypeptides with cytokine-like activities, useful  
in diagnosis and gene therapy.

PS Claim 20; Page 3253-3254, 6221pp; English.

XX  
CC The invention relates to polynucleotides (AAK51456-AAK51435) and the  
CC encoded polypeptides (AAW80323-AAW80302) that exhibit activity elating to  
CC cytokine, cell proliferation or cell differentiation or which may induce  
CC production of other cytokines in other cell populations. The  
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or  
CC peptide therapy. The polypeptides have various cytokine-like activities,  
CC e.g. stem cell growth factor activity, haematopoiesis regulating  
CC activity, tissue growth factor activity, immunomodulatory activity and  
CC activity/inhibit activity and may be useful in the diagnosis and/or  
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and  
CC inflammation. Note: Records for SEQ ID NO 2110 (AAK52581), 2111  
CC (AAK52582) and 3666 (AAW80020) are omitted as the relevant pages from the  
CC sequence listing were missing at the time of publication

XX  
SQ Sequence 526 AA;

Query Match 35.3%; Score 1282; DB 4; Length 526;

Best Local Similarity 45.0%; Pred. No. 3.6e-71;

Matches 301; Conservative 59; Mismatches 153; Indels 156; Gaps 30;

QY 1 MASTDYSTYSQAQAQGYSAVTAOPTQGYA-OTTAAYGQSSYGTGQPTDVSTYTAQTGA 59  
DB 1 MANDNTY-----QATQSYGAYPTQPGGYGQSSQPYGQSSYSGYSTD-----TS 48  
QY 60 TYGQTAATATSYGQP--TGTTPTAPQAYSQPVQYG--TGAAYDTTATVTTTQASVAAQSA 117  
DB 49 GYGQSSY--SSYGSQNTGYGTQSTP-----CGYSTGTGYG-----SSQSSQSS 90  
QY 118 YGTQPAVPAVGGQPAATAPTRPDGKPTETSQPSQSTGYNQPSLGYGSNTSYQVPG 177  
DB 91 YGQSSSYPGGQPA-----PSSYSGSYGSS--QSSSYGQPPQ--SG 128  
QY 178 SYMOPPTAPPSYPTSYSTQPTSYDQSSYQOONTYQPSYQOQSSYQOQSSYQOQPP 237  
DB 129 SYGQP-----SYG-----QOQSYGQOQSTNPQYQOQNYSSGCGGCGG 172  
QY 238 TSYPTQTSYSQAPSOYQSSSYGQSSSPRDHPSSMGVYQESGSGFSGPGENRSMGP 297  
DB 173 GG-----GNYGQOQSSMSGSGGSGGQYGNQDQSGGSGGSGYQQDRG---GRGRGSG- 222  
QY 298 DNRGRGGRGDR--GGMSSRGGRGGRGGMSSAG--ERGGYKPGGPMDEGPDLDLPVD 353  
DB 223 GGGGGGGGGYNRSSGGEPRGRGGRGGMSSDRGFMKFGGPRDQSRHD-----S 277  
QY 354 PDESDNSAIYVQGLNDSTLDDLPDAPFKCGVYKNNKRTGQPMIHYLDKETGKPKDA 413  
DB 278 EQDNDNDNTTFVQGLGENTVIESYADYFKQIGIITKTKTGQPMINIYDRETKLGEA 337  
QY 414 TVSYEDPEPTAKAAVEWFDKDFQGSKLKVLARKKPPMNSMRGLPPREGGMPPLRG 473  
DB 338 TVSFDDPPSAKAAIDWDFGKEFGNPIKVSFAIRBADFN--RGGNGRGRG-----RG 390  
QY 474 PGPPGPGGPMGMGMRGGRGGRGPPPRGSGRGNBEGGNYQHRAGDMQCPNPGCANON 533  
DB 391 PMRGGGYGG-----GSGGGGGRGPPSG-----GGGGGQORAGDMRCFNPTECMMN 437  
QY 534 FAARTCNOCAPKEPGLPPPPPPGDRGRGPGGMR--GGRGLMDRGRGPGGMFRG-- 590  
DB 438 FSHRAETCNOCKAKRPDQ-----FG-----GPGGSHMGNTYGDRRGRGGRGYDRGY 484  
QY 591 -GGGDRGGRGGR--GMDRGFGGGRGGRGPGPLMEQMGRRGGRGGRGGRG--KGEH 647  
DB 485 RGRGDRGGRGGRGGRGGRG-----GPGKMDSRGEBH 517  
QY 648 RQERDRPY 656  
DB 518 RODRRERPY 526

Search completed: February 18, 2005, 15:09:03  
Job time : 175 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 18, 2005, 14:59:05 ; Search time 47 Seconds  
(without alignments)  
1041.910 Million cell updates/sec

Title: US-10-791-017A-2

Perfect score: 3633  
Sequence: 1 MASTDYSTRSQAAAGGYSIA.....GGFGKADKGEHQRDRPY 656

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-Processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents AA:\*  
1: /cgn2\_6/prodata/1/1aa/5A\_COMB.pep:\*  
2: /cgn2\_6/prodata/1/1aa/5B\_COMB.pep:\*  
3: /cgn2\_6/prodata/1/1aa/5A\_COMB.pep:\*  
4: /cgn2\_6/prodata/1/1aa/5B\_COMB.pep:\*  
5: /cgn2\_6/prodata/1/1aa/5A\_COMB.pep:\*  
6: /cgn2\_6/prodata/1/1aa/5B\_COMB.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query	Match Length	ID	Description
1	3633	100.0	656	2 US-08-343-443B-2	Sequence 2, Appl1
2	3633	100.0	656	3 US-09-214-564A-4	Sequence 4, Appl1
3	3633	100.0	656	4 US-09-538-092-1250	Sequence 1250, Ap
4	3178.5	87.5	591	4 US-09-949-016-10914	Sequence 10914, A
5	3178.5	87.5	591	4 US-09-949-016-10915	Sequence 10915, A
6	1420	39.1	362	1 US-08-437-027-21	Sequence 21, Appl
7	1420	39.1	365	1 US-08-437-027-20	Sequence 20, Appl
8	1337	36.8	306	4 US-09-949-016-7825	Sequence 7825, Ap
9	1282	35.3	526	4 US-09-538-092-1080	Sequence 1080, Ap
10	484	13.3	295	4 US-08-343-443B-107	Sequence 107, App
11	470.5	13.0	462	4 US-09-919-039-324	Sequence 324, App
12	453	12.5	86	2 US-08-343-443B-7	Sequence 7, Appl1
13	376.5	10.4	1078	3 US-08-963-825-21	Sequence 21, Appl
14	376.5	10.4	1078	3 US-09-500-811-21	Sequence 21, Appl
15	376.5	10.4	1078	3 US-09-570-573-21	Sequence 21, Appl
16	376.5	10.4	1078	3 US-09-548-608-21	Sequence 21, Appl
17	373	10.3	1057	3 US-08-931-820-4	Sequence 4, Appl1
18	344	9.5	745	2 US-09-010-928B-28	Sequence 28, Appl1
19	344	9.5	870	2 US-09-010-928B-4	Sequence 4, Appl1
20	339.5	9.3	907	2 US-09-010-928B-2	Sequence 2, Appl1
21	338.5	9.3	1065	1 US-08-642-255-72	Sequence 72, Appl
22	337.5	9.3	1461	4 US-09-585-887-9	Sequence 9, Appl1
23	337.5	9.3	1461	4 US-09-289-578-9	Sequence 9, Appl1
24	335.5	9.2	1418	3 US-08-963-825-20	Sequence 20, Appl
25	335.5	9.2	1418	3 US-09-010-999-1	Sequence 1, Appl1
26	335.5	9.2	1418	3 US-09-500-811-20	Sequence 20, Appl
27	335.5	9.2	1418	3 US-09-570-573-20	Sequence 20, Appl

28	335.5	9.2	1418	3 US-09-548-608-20	Sequence 20, Appl
29	334.5	9.2	633	1 US-08-642-255-73	Sequence 73, Appl
30	334	9.2	1739	4 US-09-795-061-2	Sequence 2, Appl1
31	333.5	9.2	714	3 US-08-556-978B-61	Sequence 61, Appl
32	333.5	9.2	714	3 US-09-247-806-10	Sequence 10, Appl
33	333	9.2	1464	4 US-09-331-347C-21	Sequence 21, Appl
34	331	9.1	1057	3 US-08-931-820-1	Sequence 1, Appl1
35	331	9.1	1745	4 US-09-795-061-4	Sequence 4, Appl1
36	330	9.1	1806	4 US-09-919-497-56	Sequence 56, Appl
37	329	9.1	822	3 US-09-219-849-49	Sequence 49, Appl
38	327.5	9.0	1060	3 US-08-931-820-3	Sequence 3, Appl1
39	327	9.0	492	4 US-08-468-996-12	Sequence 12, Appl
40	326.5	9.0	1017	4 US-08-468-996-10	Sequence 10, Appl
41	324.5	8.9	1341	3 US-08-963-825-18	Sequence 18, Appl
42	324.5	8.9	1341	3 US-09-500-811-18	Sequence 18, Appl
43	324.5	8.9	1341	3 US-09-570-573-18	Sequence 18, Appl
44	324.5	8.9	1341	3 US-09-548-608-18	Sequence 18, Appl
45	322	8.9	1179	4 US-09-949-016-708B	Sequence 708B, Ap

#### ALIGNMENTS

RESULT 1  
US-08-343-443B-2  
Sequence 2, Application US/08343443B  
Patent No. 5968734  
GENERAL INFORMATION:  
APPLICANT: Auria, Alain  
APPLICANT: Delatree, Olivier  
APPLICANT: Desmarte, Chantal  
APPLICANT: Meiot, Thomas  
APPLICANT: Peter, Martine  
APPLICANT: Ploougaestel, Beatrice  
APPLICANT: Thomas, Gilles  
APPLICANT: Zucman, Jessica  
TITLE OF INVENTION: NUCLEIC ACID CORRESPONDING TO A GENE OF  
TITLE OF INVENTION: CHROMOSOME 22 INVOLVED IN RECURRENT CHROMOSOMAL  
TITLE OF INVENTION: TRANSLATIONS ASSOCIATED WITH THE DEVELOPMENT OF CANCEROUS  
TITLE OF INVENTION: TUMORS, AND NUCLEIC ACIDS OF FUSION RESULTING FROM SAID  
TITLE OF INVENTION: TRANSLOCATIONS  
NUMBER OF SEQUENCES: 129  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Weiser & Associates  
STREET: 230 South Fifteenth Street  
CITY: Philadelphia  
STATE: PA  
COUNTRY: USA  
ZIP: 19102  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: AEDIT 1.0 DOS text editor  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/343,443B  
FILING DATE: 18-NOV-1994  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/FR93/00494  
FILING DATE: 19-MAY-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: FR 92/06123  
FILING DATE: 20-MAY-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Weiser, Gerard J.  
REGISTRATION NUMBER: 19,763  
REFERENCE/DOCKET NUMBER: 989,6121P  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 215-875-8363  
TELEFAX: 215-875-8394  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:

LENGTH: 656 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 US-08-343-443B-2

Query Match 100.0%; Score 3633; DB 2; Length 656;  
 Best Local Similarity 100.0%; Pred. No. 8.4e-243;  
 Matches 656; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 MASTDSTYSQAQAAGGSAVTAQPTGVAQTTOAYGQSSYGYGQPTDVSYTAQQTAT 60
DB 1 MASTDSTYSQAQAAGGSAVTAQPTGVAQTTOAYGQSSYGYGQPTDVSYTAQQTAT 60
QY 61 YGQTAATSYGQPTGPTTTPAQAAYSQPVQGYGTGAYDTTATVTTTQASVAAQSAVGT 120
DB 61 YGQTAATSYGQPTGPTTTPAQAAYSQPVQGYGTGAYDTTATVTTTQASVAAQSAVGT 120
QY 121 QPAVPAVGGQPPATATPRTPDGKPTETISQPSSTGYNQPSLIGYQSNYSYQVPGSY 180
DB 121 QPAVPAVGGQPPATATPRTPDGKPTETISQPSSTGYNQPSLIGYQSNYSYQVPGSY 180
QY 181 MQPVTAPESTPPTSYSTQPTSYDQSSYQONTYGPSSYGOQSSYGOQSSYGOQPTSY 240
DB 181 MQPVTAPESTPPTSYSTQPTSYDQSSYQONTYGPSSYGOQSSYGOQSSYGOQPTSY 240
QY 241 PPTGTSYQAPSPQYSSQSSSYGOQSSYQONTYGPSSYGOQSSYGOQSSYGOQPTSY 300
DB 241 PPTGTSYQAPSPQYSSQSSSYGOQSSYQONTYGPSSYGOQSSYGOQSSYGOQPTSY 300
QY 301 GRRGGRFDRGMSRGRGGRGMSAGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGR 360
DB 301 GRRGGRFDRGMSRGRGGRGMSAGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGR 360
QY 361 SAIVYGLNDSVTLDDLAFKQCGVVKMKRTGQPMIHIYLDKTEKPKGDAVSYEDP 420
DB 361 SAIVYGLNDSVTLDDLAFKQCGVVKMKRTGQPMIHIYLDKTEKPKGDAVSYEDP 420
QY 421 PTKAAVEMFDGKDFGSKLKVSLARKKPPNMSMRGGLPREGGMPPLRGSGGPGGP 480
DB 421 PTKAAVEMFDGKDFGSKLKVSLARKKPPNMSMRGGLPREGGMPPLRGSGGPGGP 480
QY 481 GGMWGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGR 540
DB 481 GGMWGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGR 540
QY 541 NCKKAPKEGFLPPPPPPGDRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGR 600
DB 541 NCKKAPKEGFLPPPPPPGDRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGR 600
QY 601 GGRGMDRGFGGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGR 656
DB 601 GGRGMDRGFGGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGR 656

```

## RESULT 2

US-09-214-564A-4  
 ; Sequence 4, Application US/09214564A

Patent No. 6150515  
 GENERAL INFORMATION:  
 APPLICANT: Sharp, Phillip A.  
 APPLICANT: Zhou, Qiang  
 TITLE OF INVENTION: TAT-SF: Cofactor For Stimulation Of Transcriptional  
 TITLE OF INVENTION: Elongation By HIV-1 TAT  
 FILE REFERENCE: M0656/7042  
 CURRENT APPLICATION NUMBER: US/09/214,564A  
 PRIOR FILING DATE: 1999-08-18  
 PRIOR FILING DATE: 1996-07-03  
 PRIOR FILING DATE: 1996-07-03  
 PRIOR FILING DATE: 1996-07-03  
 PRIOR FILING DATE: 1996-12-13  
 PRIOR APPLICATION NUMBER: PCT/US97/11713  
 PRIOR FILING DATE: 1997-07-03

NUMBER OF SEQ ID NOS: 5  
 SOFTWARE: FastSeq for Windows Version 3.0  
 SEQ ID NO 4  
 LENGTH: 656  
 TYPE: PRT  
 ORGANISM: Homo sapiens  
 US-09-214-564A-4

Query Match 100.0%; Score 3633; DB 3; Length 656;  
 Best Local Similarity 100.0%; Pred. No. 8.4e-243;  
 Matches 656; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 MASTDSTYSQAQAAGGSAVTAQPTGVAQTTOAYGQSSYGYGQPTDVSYTAQQTAT 60
DB 1 MASTDSTYSQAQAAGGSAVTAQPTGVAQTTOAYGQSSYGYGQPTDVSYTAQQTAT 60
QY 61 YGQTAATSYGQPTGPTTTPAQAAYSQPVQGYGTGAYDTTATVTTTQASVAAQSAVGT 120
DB 61 YGQTAATSYGQPTGPTTTPAQAAYSQPVQGYGTGAYDTTATVTTTQASVAAQSAVGT 120
QY 121 QPAVPAVGGQPPATATPRTPDGKPTETISQPSSTGYNQPSLIGYQSNYSYQVPGSY 180
DB 121 QPAVPAVGGQPPATATPRTPDGKPTETISQPSSTGYNQPSLIGYQSNYSYQVPGSY 180
QY 181 MQPVTAPESTPPTSYSTQPTSYDQSSYQONTYGPSSYGOQSSYGOQSSYGOQPTSY 240
DB 181 MQPVTAPESTPPTSYSTQPTSYDQSSYQONTYGPSSYGOQSSYGOQSSYGOQPTSY 240
QY 241 PPTGTSYQAPSPQYSSQSSSYGOQSSYQONTYGPSSYGOQSSYGOQSSYGOQPTSY 300
DB 241 PPTGTSYQAPSPQYSSQSSSYGOQSSYQONTYGPSSYGOQSSYGOQSSYGOQPTSY 300
QY 301 GRRGGRFDRGMSRGRGGRGMSAGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGR 360
DB 301 GRRGGRFDRGMSRGRGGRGMSAGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGR 360
QY 361 SAIVYGLNDSVTLDDLAFKQCGVVKMKRTGQPMIHIYLDKTEKPKGDAVSYEDP 420
DB 361 SAIVYGLNDSVTLDDLAFKQCGVVKMKRTGQPMIHIYLDKTEKPKGDAVSYEDP 420
QY 421 PTKAAVEMFDGKDFGSKLKVSLARKKPPNMSMRGGLPREGGMPPLRGSGGPGGP 480
DB 421 PTKAAVEMFDGKDFGSKLKVSLARKKPPNMSMRGGLPREGGMPPLRGSGGPGGP 480
QY 481 GGMWGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGR 540
DB 481 GGMWGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGR 540
QY 541 NCKKAPKEGFLPPPPPPGDRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGR 600
DB 541 NCKKAPKEGFLPPPPPPGDRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGR 600
QY 601 GGRGMDRGFGGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGR 656
DB 601 GGRGMDRGFGGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGR 656

```

## RESULT 3

US-09-538-092-1250  
 ; Sequence 1250, Application US/09538092

Patent No. 6753314  
 GENERAL INFORMATION:  
 APPLICANT: Glot, Loic  
 APPLICANT: Manfield, Traci A.  
 TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same  
 FILE REFERENCE: 15966-542  
 CURRENT APPLICATION NUMBER: US/09/538,092  
 PRIOR FILING DATE: 2000-03-29  
 PRIOR FILING DATE: 1999-04-01  
 PRIOR FILING DATE: 1999-04-01  
 PRIOR FILING DATE: 2000-02-01  
 NUMBER OF SEQ ID NOS: 1387



```
/ SOFTWARE: CuraPatSeqFormatter Version 0.9
/ SEQ ID NO 1250
/ LENGTH: 656
/ TYPE: PRT
/ ORGANISM: Homo sapiens
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: (0)...(0)
/ OTHER INFORMATION: Polypeptide Accession Number Q01844
US-09-538-092-1250
```

```
Query Match      100.0%; Score 3633; DB 4; Length 656;
Best Local Similarity 100.0%; Pred. No. 8.4e-243;
Matches 656; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 MASTDSTVSYQAQAAGQSAVTAQPTQGYAQTTOAYGQGSYGTGYPPTDVSYTOAQTAT 60
DB 1 MASTDSTVSYQAQAAGQSAVTAQPTQGYAQTTOAYGQGSYGTGYPPTDVSYTOAQTAT 60
QY 61 YGQTAATSYGQPTGTTTPTAPQAYSQPVQGYGTGAYDTTATVTTTQASVAAQSAYGT 120
DB 61 YGQTAATSYGQPTGTTTPTAPQAYSQPVQGYGTGAYDTTATVTTTQASVAAQSAYGT 120
QY 121 QPAPYAGQOPATATPRTPODGNKPTETSQPOSTGTGYNPISLGYGSNTSYPOVPSYP 180
DB 121 QPAPYAGQOPATATPRTPODGNKPTETSQPOSTGTGYNPISLGYGSNTSYPOVPSYP 180
QY 181 MQPVTAPPSYPTSYSTQPTSYDQSSYQONTYGGPSSYGOQSSYGOQSSYGOQPTSY 240
DB 181 MQPVTAPPSYPTSYSTQPTSYDQSSYQONTYGGPSSYGOQSSYGOQSSYGOQPTSY 240
QY 241 PPTGTSYQAAPSOYSQOQSSYGOQSSYFRQDHPSSMGVYGOESGSGFSGGENRSMGSDNR 300
DB 241 PPTGTSYQAAPSOYSQOQSSYGOQSSYFRQDHPSSMGVYGOESGSGFSGGENRSMGSDNR 300
QY 301 GRGRGGRFDRGMSRGRGGRGMSAGRGGRFNRKPGPMDDEGDDLDLGPVPDDESDN 360
DB 301 GRGRGGRFDRGMSRGRGGRGMSAGRGGRFNRKPGPMDDEGDDLDLGPVPDDESDN 360
QY 361 SAIVYQGLNDSVTLDDLADFFKQCGVVKMKRTGQPMIHLYLDKETGPKGDAVSYEDP 420
DB 361 SAIVYQGLNDSVTLDDLADFFKQCGVVKMKRTGQPMIHLYLDKETGPKGDAVSYEDP 420
QY 421 PTKAAAVENFDGDFQSSKLVSLARKKPPMNSMRGGLPREBGRMPPLRGPGGPGGGR 480
DB 421 PTKAAAVENFDGDFQSSKLVSLARKKPPMNSMRGGLPREBGRMPPLRGPGGPGGGR 480
QY 481 GGPMSGRGGRGGRGGRGPRGPRGSRGNPSGGGVVQHRADWQCPNPGCGNONPAMRTEC 540
DB 481 GGPMSGRGGRGGRGGRGPRGPRGSRGNPSGGGVVQHRADWQCPNPGCGNONPAMRTEC 540
QY 541 NCKAKPKEGFLPPLPPPPGCDRGGRGPGGMRGGRGGLMDRGPGGGMFRGGRGGRGGR 600
DB 541 NCKAKPKEGFLPPLPPPPGCDRGGRGPGGMRGGRGGLMDRGPGGGMFRGGRGGRGGR 600
QY 601 GGGGMDRGGRGGRGGRGPGPPLMEQMGRRRGGRGPGKMDKGHRQRRDRPY 656
DB 601 GGGGMDRGGRGGRGGRGPGPPLMEQMGRRRGGRGPGKMDKGHRQRRDRPY 656
```

```
RESULT 4
US-09-949-016-10914
/ Sequence 10914, Application US/09949016
/ Patent No. 6812339
/ GENERAL INFORMATION:
/ APPLICANT: VENTER, J. Craig et al.
/ TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
/ FILE REFERENCE: C1001307
/ CURRENT APPLICATION NUMBER: US/09/949,016
/ PRIOR APPLICATION NUMBER: 2000-04-14
/ PRIOR FILING DATE: 2000-10-20
```

```
/ PRIOR APPLICATION NUMBER: 60/237,768
/ PRIOR FILING DATE: 2000-10-03
/ PRIOR APPLICATION NUMBER: 60/231,498
/ PRIOR FILING DATE: 2000-09-08
/ NUMBER OF SEQ ID NOS: 207012
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 10914
/ LENGTH: 591
/ TYPE: PRT
/ ORGANISM: Human
US-09-949-016-10914
```

```
Query Match      87.5%; Score 3178.5; DB 4; Length 591;
Best Local Similarity 88.7%; Pred. No. 1.7e-211;
Matches 582; Conservative 0; Mismatches 1; Indels 73; Gaps 1;
```

```
QY 1 MASTDSTVSYQAQAAGQSAVTAQPTQGYAQTTOAYGQGSYGTGYPPTDVSYTOAQTAT 60
DB 9 MASTDSTVSYQAQAAGQSAVTAQPTQGYAQTTOAYGQGSYGTGYPPTDVSYTOAQTAT 68
QY 61 YGQTAATSYGQPTGTTTPTAPQAYSQPVQGYGTGAYDTTATVTTTQASVAAQSAYGT 120
DB 69 YGQTAATSYGQPTGTTTPTAPQAYSQPVQGYGTGAYDTTATVTTTQASVAAQSAYGT 128
QY 121 QPAPYAGQOPATATPRTPODGNKPTETSQPOSTGTGYNPISLGYGSNTSYPOVPSYP 180
DB 129 QPAPYAGQOPATATPRTPODGNKPTETSQPOSTGTGYNPISLGYGSNTSYPOVPSYP 188
QY 181 MQPVTAPPSYPTSYSTQPTSYDQSSYQONTYGGPSSYGOQSSYGOQSSYGOQPTSY 240
DB 189 MQPVTAPPSYPTSYSTQPTSYDQSSYQONTYGGPSSYGOQSSYGOQSSYGOQPTSY 248
QY 241 PPTGTSYQAAPSOYSQOQSSYGOQSSYFRQDHPSSMGVYGOESGSGFSGGENRSMGSDNR 300
DB 249 PPTGTSYQAAPSOYSQOQSSYGOQSSYFRQDHPSSMGVYGOESGSGFSGGENRSMGSDNR 300
QY 301 GRGRGGRFDRGMSRGRGGRGMSAGRGGRFNRKPGPMDDEGDDLDLGPVPDDESDN 360
DB 274 -----PMDDEGDDLDLGPVPDDESDN 295
QY 361 SAIVYQGLNDSVTLDDLADFFKQCGVVKMKRTGQPMIHLYLDKETGPKGDAVSYEDP 420
DB 296 SAIVYQGLNDSVTLDDLADFFKQCGVVKMKRTGQPMIHLYLDKETGPKGDAVSYEDP 355
QY 421 PTKAAAVENFDGDFQSSKLVSLARKKPPMNSMRGGLPREBGRMPPLRGPGGPGGGR 480
DB 356 PTKAAAVENFDGDFQSSKLVSLARKKPPMNSMRGGLPREBGRMPPLRGPGGPGGGR 415
QY 481 GGPMSGRGGRGGRGGRGPRGPRGSRGNPSGGGVVQHRADWQCPNPGCGNONPAMRTEC 540
DB 416 GGPMSGRGGRGGRGGRGPRGPRGSRGNPSGGGVVQHRADWQCPNPGCGNONPAMRTEC 475
QY 541 NCKAKPKEGFLPPLPPPPGCDRGGRGPGGMRGGRGGLMDRGPGGGMFRGGRGGRGGR 600
DB 476 NCKAKPKEGFLPPLPPPPGCDRGGRGPGGMRGGRGGLMDRGPGGGMFRGGRGGRGGR 535
QY 601 GGGGMDRGGRGGRGGRGPGPPLMEQMGRRRGGRGPGKMDKGHRQRRDRPY 656
DB 536 GGGGMDRGGRGGRGGRGPGPPLMEQMGRRRGGRGPGKMDKGHRQRRDRPY 591
```

```
RESULT 5
US-09-949-016-10915
/ Sequence 10915, Application US/09949016
/ Patent No. 6812339
/ GENERAL INFORMATION:
/ APPLICANT: VENTER, J. Craig et al.
/ TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
/ FILE REFERENCE: C1001307
/ CURRENT APPLICATION NUMBER: US/09/949,016
/ PRIOR APPLICATION NUMBER: 2000-04-14
/ PRIOR FILING DATE: 2000-10-20
```

PRIOR FILING DATE: 2000-10-20  
PRIOR APPLICATION NUMBER: 60/237,768  
PRIOR FILING DATE: 2000-10-03  
PRIOR APPLICATION NUMBER: 60/231,498  
PRIOR FILING DATE: 2000-09-08  
NUMBER OF SEQ ID NOS: 207012  
SOFTWARE: Faeseq for Windows Version 4.0  
SEQ ID NO 10915  
LENGTH: 591  
TYPE: PRT  
ORGANISM: Human  
US-09-949-016-10915

Query Match 87.5%; Score 3178.5; DB 4; Length 591;  
Best Local Similarity 88.7%; Pred. No. 1.7e-211; Indels 73; Gaps 1;  
Matches 582; Conservative 0; Mismatches 1;

QY 1 MASTDVSYSQAQAAGGSAATTAQPTGVAQTTOAYGQSSYGTGPTDVSYTOAQTAT 60  
DB 9 MASTDVSYSQAQAAGGSAATTAQPTGVAQTTOAYGQSSYGTGPTDVSYTOAQTAT 68  
QY 61 YGCTAATATSGPPTGTTTTPAQAQSAQPVQGTGAYDTTATVTTTQASAAQSAYGT 120  
DB 69 YGCTAATATSGPPTGTTTTPAQAQSAQPVQGTGAYDTTATVTTTQASAAQSAYGT 128  
QY 121 QPAYPAVGOQPAATAPTRPDGKPKPTTSQPOSTGSGNPSLIGYGQSNYSYQVPGSY 180  
DB 129 QPAYPAVGOQPAATAPTRPDGKPKPTTSQPOSTGSGNPSLIGYGQSNYSYQVPGSY 188  
QY 181 MOPVTAPPSYPTSTSTQPTSTYDQSSYQONTYGOBSYGOQSSYGOQSSYGOQPTSY 240  
DB 189 MOPVTAPPSYPTSTSTQPTSTYDQSSYQONTYGOBSYGOQSSYGOQSSYGOQPTSY 248  
QY 241 PPTGTSYQAAPSOYSSQSSSYGQSSFRQDHPBSMGVYGGSSGSGFSGPGENRMSGPDNR 300  
DB 249 PPTGTSYQAAPSOYSSQSSSYGQOR----- 273  
QY 301 GRGRGGRGDRGMRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGR 360  
DB 274 -----PMDGPDLDLGGPVDPDESDN 235  
QY 361 SAIVYGLANDSVTLDDLADLPKOCGVVKNMKTGPWMIHLYLXETGKPGDAATSYEDP 420  
DB 296 SAIVYGLANDSVTLDDLADLPKOCGVVKNMKTGPWMIHLYLXETGKPGDAATSYEDP 355  
QY 421 PTKAAVEMWEDGDFQSSKLVSLARKKPPMNSMRGGLPRERGRMPPLRGGPGGGGR 480  
DB 356 PTKAAVEMWEDGDFQSSKLVSLARKKPPMNSMRGGLPRERGRMPPLRGGPGGGGR 415  
QY 481 GGMGMRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGR 540  
DB 416 GGMGMRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGR 475  
QY 541 NCKAPKPEGLFPPEPPGDRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGR 600  
DB 476 NCKAPKPEGLFPPEPPGDRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGR 535  
QY 601 GGRGMRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGR 656  
DB 536 GGRGMRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGR 591

RESULT 6  
US-08-437-027-21  
Sequence 21, Application US/08437027  
Patent No. 5670317  
GENERAL INFORMATION:  
APPLICANT: Landanyi, Marc  
TITLE OF INVENTION: A DIAGNOSTIC TEST FOR TEST FOR THE DESMOPLASTIC  
NUMBER OF SEQUENCES: 21  
CORRESPONDENCE ADDRESS: 21

TITLE OF INVENTION: A DIAGNOSTIC TEST FOR TEST FOR THE DESMOPLASTIC  
NUMBER OF SEQUENCES: 21  
CORRESPONDENCE ADDRESS: 21

ADDRESSER: Cooper & Dunham LLP  
STREET: 1185 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/437,027  
FILING DATE:  
CLASSIFICATION: 536  
ATTORNEY/AGENT INFORMATION:  
NAME: White, John P  
REGISTRATION NUMBER: 28,678  
REFERENCE/DOCKET NUMBER: 46416/JPW/CCA  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-278-0400  
TELEFAX: 212-391-0525  
INFORMATION FOR SEQ ID NO: 21:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 362 amino acids  
TYPE: amino acid  
STRANDEDNESS: unknown  
TOPOLOGY: unknown  
MOLECULE TYPE: protein  
US-08-437-027-21

Query Match 39.1%; Score 1420; DB 1; Length 362;  
Best Local Similarity 100.0%; Pred. No. 2.2e-90; Indels 0; Gaps 0;  
Matches 265; Conservative 0; Mismatches 0;

QY 1 MASTDVSYSQAQAAGGSAATTAQPTGVAQTTOAYGQSSYGTGPTDVSYTOAQTAT 60  
DB 1 MASTDVSYSQAQAAGGSAATTAQPTGVAQTTOAYGQSSYGTGPTDVSYTOAQTAT 60  
QY 61 YGCTAATATSGPPTGTTTTPAQAQSAQPVQGTGAYDTTATVTTTQASAAQSAYGT 120  
DB 61 YGCTAATATSGPPTGTTTTPAQAQSAQPVQGTGAYDTTATVTTTQASAAQSAYGT 120  
QY 121 QPAYPAVGOQPAATAPTRPDGKPKPTTSQPOSTGSGNPSLIGYGQSNYSYQVPGSY 180  
DB 121 QPAYPAVGOQPAATAPTRPDGKPKPTTSQPOSTGSGNPSLIGYGQSNYSYQVPGSY 180  
QY 181 MOPVTAPPSYPTSTSTQPTSTYDQSSYQONTYGOBSYGOQSSYGOQSSYGOQPTSY 240  
DB 181 MOPVTAPPSYPTSTSTQPTSTYDQSSYQONTYGOBSYGOQSSYGOQSSYGOQPTSY 240  
QY 241 PPTGTSYQAAPSOYSSQSSSYGQSSFRQDHPBSMGVYGGSSGSGFSGPGENRMSGPDNR 300  
DB 241 PPTGTSYQAAPSOYSSQSSSYGQSS----- 265

RESULT 7  
US-08-437-027-20  
Sequence 20, Application US/08437027  
Patent No. 5670317  
GENERAL INFORMATION:  
APPLICANT: Landanyi, Marc  
TITLE OF INVENTION: A DIAGNOSTIC TEST FOR TEST FOR THE DESMOPLASTIC  
NUMBER OF SEQUENCES: 21  
CORRESPONDENCE ADDRESS: 21  
ADDRESSER: Cooper & Dunham LLP  
STREET: 1185 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036

TITLE OF INVENTION: A DIAGNOSTIC TEST FOR TEST FOR THE DESMOPLASTIC  
NUMBER OF SEQUENCES: 21  
CORRESPONDENCE ADDRESS: 21



QY	178	SYMGGVLTAPSPAPSPFTSYSTQPTASVDSSVYQXNTYQPSYSYQQSSSYGQQSSYGGQPP	2.7
Db	129	SYGQPF-----SYGG-----QQSSYQQSSYMPPGCTGQGNQYNSSSGGGGGG	1.22
QY	238	TSYPTQTSISQAPSOYSQQSSSYQQSSFRODHPSSMGVTGQSSGGFSGGGRNSMGP	2.97
Db	173	GG-----GNYGQDOSMSSSGGSGGGYGQNDSSGGSGGGYGQDDRG-----GRGGSGG-	2.22
QY	298	DNRGRGGGPR--GGSMSRGGRGGRGGMGMSAG--ERGAFYKPGGPMDEGDDLGPVYD	3.53
Db	223	GGGGGGGGGYNSSSGYEPRGRGGGRGGRGCMGSGSDRGGRFKKFGPRDQSSRH-----S	2.77
QY	354	PDESDNSAIYYQGLNDVTLDDLADPFKQCGVVMKNTRTQPMIHLYLDEKTEKPKGDA	4.13
Db	278	EQNSNHNNTIFVOGLDENVTIISVADYFKQIGIITKTKKTQPMINLTDTRETEKLGEEA	3.37
QY	414	TVSEYEDPPTAKAAVEMFDGKDFQSGAKLYSLARKPPMNSMRGGLPREBGMPPLRG	4.73
Db	338	TYSFDDPPSKAALIDFQKEPSSGNPIKVSFATRPADPN--RGGNGRGGR-----RGG	3.90
QY	474	PGSPGGPGPMGRMGRGGDRGGPFRGRBGRSNPSGGGVYORADPMQCPNPGCGNON	5.33
Db	391	PMRGCGYGGG-----GSGGGGRGGPSSGG-----GGGGGQGRAGGDKCPNPTCEMNN	4.37
QY	534	FAMRTSCNCKAKPKPPGFLPPPPPPGCDRGMGGPGMR--GGRGLMDRGPGGMPFG--	5.50
Db	438	FSHMRNCNCKAKPKPPG-----PG-----GGGGSSHMGGANTGDDRRGGGGYDGGY	4.84
QY	591	-GRGGDRGGFRGR--GMDRGGFGGRRGPGGPGPGLMEQWGRGRGRGGPKMD--KGEH	6.47
Db	485	RGGGGDRGGFRGGRGGRGGDRGF-----GGKMDSRGEH	5.17
QY	648	ROERRDRPY	6.56
Db	518	RODRRRPY	5.26

US-RESULT 10  
US-08-343-443B-107  
Sequence 107, Application US/08343443B  
Patent No. 5968734  
GENERAL INFORMATION:  
APPLICANT: Auitas, Alain  
APPLICANT: Delattre, Olivier  
APPLICANT: Desmazes, Chantal  
APPLICANT: Melot, Thomas  
APPLICANT: Peter, Martine  
APPLICANT: Ploongastel, Beatrice  
APPLICANT: Thomas, Gilles  
APPLICANT: Zuchman, Vessica  
TITLE OF INVENTION: NUCLEIC ACID CORRESPONDING TO A GENE OF  
TITLE OF INVENTION: CHROMOSOME 22 INVOLVED IN RECURRENT CHROMOSOMAL  
TITLE OF INVENTION: TRANSLATIONS ASSOCIATED WITH THE DEVELOPMENT OF CANCEROUS  
TITLE OF INVENTION: TUMORS, AND NUCLEIC ACIDS OF FUSION RESULTING FROM SAID  
TITLE OF INVENTION: TRANSLATIONS  
NUMBER OF SEQUENCES: 129  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Welser & Associates  
STREET: 230 South Fifteenth Street  
CITY: Philadelphia  
STATE: PA  
COUNTRY: USA  
ZIP: 19102  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: AEDIT 1.0 DOS text editor  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/343,443B  
FILING DATE: 18-NOV-1994  
CLASSIFICATION: 514

```

PROR APPLICATION DATA:
APPLICATION NUMBER: PCT/FR93/00494
FILING DATE: 19-MAY-1993
PROR APPLICATION DATA:
APPLICATION NUMBER: FR 92/06123
FILING DATE: 20-MAY-1992
ATTORNEY/AGENT INFORMATION:
NAME: Weiser, Gerard J.
REGISTRATION NUMBER: 19,763
REFERENCE/DOCKET NUMBER: 989.6121E
TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-875-8383
TELEFAX: 215-875-8394
INFORMATION FOR SEQ ID NO: 107:
SEQUENCE CHARACTERISTICS:
LENGTH: 295 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-343-443B-107

```

Query Match	13.34;	Score 484;	DB 2;	Length 295;
Best Local Similarity	100.0%;	Pred. No. 6.7e-26;		
Matches 89;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

Qy	237	PTSPVPONGAGSQAAPSOYSSQSSSSYGQSSSPRROHPSMGMYYGGSSGFSFGGGRNSMG	296
Db	1	PTSPVPONGAGSQAAPSOYSSQSSSSYGQSSSPRROHPSMGMYYGGSSGFSFGGGRNSMG	60
Qy	297	PDNTGGRGRCGTFDRGMSRGCRGGCRGCMG	325
Db	61	PDNRGRGRCGTFDRGMSRGCRGGCRGCMG	89

```

RESULT 11
US-09-919-039-324
; Sequence 324, Application US/09919039
; Patent No. 6727066
; GENERAL INFORMATION:
; APPLICANT: Kaser, Matthew R.
; TITLE OF INVENTION: GENES EXPRESSED IN TREATED HUMAN C3A LIVER CELL CULTURES
; FILE REFERENCE: PA-0035 US
; CURRENT APPLICATION NUMBER: US/09/919,039
; CURRENT FILING DATE: 2002-09-09
; PRIOR APPLICATION NUMBER: 60/222,113
; PRIOR FILING DATE: 2000-07-28
; NUMBER OF SEQ ID NOS: 401
; SOFTWARE: PERL Program
; SEQ ID NO 324
; LENGTH: 462
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. 6727066 1813444CD1
US-09-919-039-324

```

	Query Match	13.0%;	Score 470.5;	DB 4;	Length 462;	
	Best Local Similarity	38.1%;	Pred. No. 9.7e-25;			
	Matches	134;	Conservative	25;	Mismatches 92;	Indels 101; Gaps 19
Cy	1	MASTDVSTYSSAQAAGSYASATTAQPTGGA-QTTQAVGSGSYGTGYDPTDVSYTQAQTTA	59			
Dd	1	MASNDYT-----QAATQSXGAVPTGPQGGSQSSQSPVGQASGVSGSTTD-----TS	48			
Cy	60	TTCGTATAATISGGPP-TGYTPPTAPQASQSPVGCG-TGAYDTTATATTQATASAAAGA	117			
Dd	49	GYGGSASY-SYGDSQONTNGYQSTP-----QCYGSGTSGYG-----SSQSSQSS	90			
Cy	118	YGNPAVPAYGQQQAATAFTAPRPGQGNKPKETSDPSQSSTNGNPNPLGYSQSNYSFPOYPG	177			
Dd	91	YGOQSSITFGIQQA-----DSTSGSSTGSSS--QSSSYGQPQ-SG	128			

[illegible]

```

Db      1 IYQGLNDSTLTDDLDLDFKQCGVVMKNTQGPMLHIYLDKETGCPKGDATVSYPDPFT 60
Qy      423 AKAAVEWFDGKDFQGSRLKTVSLARKK 448
        |||||
        |||||
Db      61 AKAAVEWFDGKDFQGSRLKTVSLARKK 86

```

```

US-08-343-443B-7
Sequence 7, Application US/08343443B
Patent No. 5968734
GENERAL INFORMATION:
APPLICANT: Aurias, Alain
APPLICANT: Delattre, Olivier
APPLICANT: Desmaze, Chantal
APPLICANT: Melot, Thomas
APPLICANT: Peter, Martine
APPLICANT: Ploougaetel, Beatrice
APPLICANT: Thomas, Gilles
APPLICANT: Zucman, Jessica
TITLE OF INVENTION: NUCLEIC ACID CORRESPONDING TO A GENE OF
TITLE OF INVENTION: CHROMOSOME 22 INVOLVED IN RECURRENT CHROMOSOMAL
TITLE OF INVENTION: TRANSLATIONS ASSOCIATED WITH THE DEVELOPMENT OF CANCEROUS
TITLE OF INVENTION: TUMORS, AND NUCLEIC ACIDS OF FUSION RESULTING FROM SAID
NUMBER OF SEQUENCES: 129
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Weiser & Associates
STREET: 230 South Fifteenth Street
CITY: Philadelphia
STATE: PA
COUNTRY: USA
ZIP: 19102
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: AEDIT 1.0 DOS text editor
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/343,443B
FILING DATE: 18-Nov-1994
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/FR93/00494
FILING DATE: 19-MAY-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: FR 92/06123
FILING DATE: 20-MAY-1992
ATTORNEY/AGENT INFORMATION:
NAME: Weiser, Gerard J.
REGISTRATION NUMBER: 19,763
REFERENCE/DOCKET NUMBER: 989.6121P
TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-875-8383
TELEFAX: 215-875-8394
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 86 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-08-343-443B-7
Query Match 12.5%; Score 453; DB 2; Length 86;
Best Local Similarity 100.0%; Pred. No. 2.2e-24;
Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
363 IYQGLNDSTVTLDDLADFFKQCGVVKMKRTGQPMHIYIDKSTGKPKGDAVTSIEDPT 422

```

```

RESULT 13
US-08-963-825-21
; Sequence 21, Application US/08963825
; Patent No. 6110689
; GENERAL INFORMATION:
; APPLICANT: Qvist, Per
; APPLICANT: Bonde, Martin
; TITLE OF INVENTION: A Method for Assaying Collagen Fragments
; TITLE OF INVENTION: In Body Fluids, A Test Kit and Means for Carrying Out the
; TITLE OF INVENTION: Method and Use of the Method to Diagnose the Presence of
; TITLE OF INVENTION: Disorders Associated with the Metabolism of
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Darby & Darby PC
; STREET: 805 Third Avenue
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/963,825
; FILING DATE:
; CLASSIFICATION: 436
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/187,319
; FILING DATE: 21-JAN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Gogoris, Adda C
; REGISTRATION NUMBER: 29,714
; REFERENCE/DOCKET NUMBER: 4305/08701
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-527-7700
; TELEFAX: 212-753-6237
; TELEX: 236687
; INFORMATION FOR SEQ ID NO: 21:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1078 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; IMMEDIATE SOURCE:
; CLONE: COLLAGEN ALPHA 1 (III)
US-08-963-825-21

Query Match      10.4%; Score 376.5; DB 3; Length 1078;
Beet Local Similarity 25.8%; Pred. No. Be-18;
Matches 184; Conservative 34; Mismatches 235; Indels 259; Gaps 34;

Dy 42 GTYGQPTDYSYTOAGTTATATGCTAATVATSYGPPGTGTTTAFAPAVSQPVQGYGTGAYDTT 101
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 39 GTSGHP-----GSRGSRCYGGPGRF--QGAGSPRG---PGALG-----75

Dy 102 TATVTTSASAAQSAYGVTPAYPAYGQPAAITAPTRPDGNKPKETSSQPSSTGCYNOP 161
    :| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:|
Db 76 -----PSGPA-GKDGESGRPGRPDRGLP-----GPPAIKGPAIGP 110

Dy 162 SL-----GYGSNTSYRQ-----VPGST----PWQPTAPPSYPTTSYSTSQ 199
    :| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:|

```

```

Db 111 GPPKMGKRGHFGDNGEKETGARGLKGENGFLPGENGAPGPMGRAP----- 158
Qy 200 PTSDYSSSYQNTYQPPSSYGOSSSYGOQPPPTYPQTSYSAQPSQSS 259
Db 159 -----GERKRPGLPGAAGARGSDQPPPG-PPGTAGFPSPG-----AKG 204
Qy 260 SYGOSSFRQDHPSSMGVYQ--ESG--GFSG-----PGENRSMGPDNRG----- 301
Db 205 EVGPAGS-----PGSNAPGQRGEPGPGHAGAGPPGPGTNGSPGKEMBPAGI 259
Qy 302 -----RGRGFDGMSRGRGGRGMSAGRGGPNKPGPMDEGPD 346
Db 260 PGLMGARPPGAPAGAPGL-RGAGBPGKNGAKGPRGRGEGAGIPGVGAKGEDG 318
Qy 347 DLGPVDPDSDNSALTYVQGLSDVTLDDLADFFQCCGVVKNKRTGPMHIYDKET 406
Db 319 KDSPGPGPANGLPGLAAGERGALGS----- 343
Qy 407 GKPKGATVSEYDEPTAKAAVEMFDKDFQSKLVSLARKKPPMNSMRG--GLPPREG 464
Db 344 RGAAGPNGIPGKGPAGE-----RGAGPAGPAGAGPGRDGV 382
Qy 465 GMPPLRGPGGPGPG--GPMGRWG--GRGDRGFPFPPRGSRG--NPSGG 512
Db 383 PGSPGMRGMPGSPGSGDKKPGPGSGESGRPPPG--PSGPRQPGVMGFPBGKN 439
Qy 513 GNVQHRAGDMQCPNGCCGNQNFAMRTCCNCKAPKPGFLPPPP-----PGDRGRGP 568
Db 440 DGAPKNGRGGPG--GPGQG-----PGKNGEYQPGPPGPTGDKDTGP 487
Qy 569 GGNRG-----GRGGLMDR-----GGPGMFRGGRGDRG-----GF 599
Db 488 RGPQGLQGLPTGTPPEENKPGEPKGEAAGAPAGAGKADGAPBERPPGLAAGPGL 547
Qy 600 RGRGMDRGFGGRRGPGPG--PLMEOWGRRGRGPG--KMDXGE 646
Db 548 RCGAG-PPEGEGKGAAGPPGPAAGTGLQGMFERGGLSGFPGKDXGE 598

```

# RESULT 14

US-09-500-811-21

Sequence 21, Application US/09500811

Patent No. 6323314

GENERAL INFORMATION:

APPLICANT: Oviat, Per

APPLICANT: Bonde, Martin

TITLE OF INVENTION: A Method for Assaying Collagen Fragments

TITLE OF INVENTION: In Body Fluids, A Test Kit and Means for Carrying Out the

TITLE OF INVENTION: Method and Use of the Method to Diagnose the Presence of

NUMBER OF SEQUENCES: 21

CORRESPONDENCE ADDRESS:

ADDRESSEE: Darby & Darby PC

STREET: 805 Third Avenue

CITY: New York

STATE: New York

COUNTRY: USA

ZIP: 10022

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/500,811

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/187,319

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Gogoris, Adda C

REGISTRATION NUMBER: 29,714

```

REFERENCE/DOCKET NUMBER: 4305/08701
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-527-7700
TELEFAX: 212-753-6237
TELEX: 236687
INFORMATION FOR SEQ ID NO: 21:
SEQUENCE CHARACTERISTICS:
LENGTH: 1078 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
IMMEDIATE SOURCE:
CLONE: COLLAGEN ALPHA 1 (III)
US-09-500-811-21

```

Query Match 10.4%; Score 376.5; DB 3; Length 1078;

Best Local Similarity 25.8%; Pred. No. 8e-18;

Matches 184; Conservative 34; Mismatches 235; Indels 259; Gaps 34;

```

Qy 42 GTTGPTDVSYQAOQTATYATYATSYGQPPPTGTTTAAQAYSQPYGTGAYDTT 101
Db 39 GTSNP-----GSPGSPYQGPGER--GAGSGPG--PGALG----- 75
Qy 102 TATVTTQASVAAQSAVGTQPAYPAYGQQAATAPTRPDGKPTETSGPQSGTVNP 161
Db 76 -----PSGPA-GKDGSGRPPRDRGLP--GPGIKGAGLP 110
Qy 162 SL-----GYQSNYSYQ-----VPSY--PMQVTPAPSYPTYSSTQ 199
Db 111 GPPKMGKRGHFGDNGEKETGARGLKGENGFLPGENGAPGPMGRAP----- 158
Qy 200 PTSDYSSSYQNTYQPPSSYGOSSSYGOQPPPTYPQTSYSAQPSQSS 259
Db 159 -----GERKRPGLPGAAGARGSDQPPPG-PPGTAGFPSPG-----AKG 204
Qy 260 SYGOSSFRQDHPSSMGVYQ--ESG--GFSG-----PGENRSMGPDNRG----- 301
Db 205 EVGPAGS-----PGSNAPGQRGEPGPGHAGAGPPGPGTNGSPGKEMBPAGI 259
Qy 302 -----RGRGFDGMSRGRGGRGMSAGRGGPNKPGPMDEGPD 346
Db 260 PGLMGARPPGAPAGAPGL-RGAGBPGKNGAKGPRGRGEGAGIPGVGAKGEDG 318
Qy 347 DLGPVDPDSDNSALTYVQGLSDVTLDDLADFFQCCGVVKNKRTGPMHIYDKET 406
Db 319 KDSPGPGPANGLPGLAAGERGALGS----- 343
Qy 407 GKPKGATVSEYDEPTAKAAVEMFDKDFQSKLVSLARKKPPMNSMRG--GLPPREG 464
Db 344 RGAAGPNGIPGKGPAGE-----RGAGPAGPAGAGPGRDGV 382
Qy 465 GMPPLRGPGGPGPG--GPMGRWG--GRGDRGFPFPPRGSRG--NPSGG 512
Db 383 PGSPGMRGMPGSPGSGDKKPGPGSGESGRPPPG--PSGPRQPGVMGFPBGKN 439
Qy 513 GNVQHRAGDMQCPNGCCGNQNFAMRTCCNCKAPKPGFLPPPP-----PGDRGRGP 568
Db 440 DGAPKNGRGGPG--GPGQG-----PGKNGEYQPGPPGPTGDKDTGP 487
Qy 569 GGNRG-----GRGGLMDR-----GGPGMFRGGRGDRG-----GF 599
Db 488 RGPQGLQGLPTGTPPEENKPGEPKGEAAGAPAGAGKADGAPBERPPGLAAGPGL 547
Qy 600 RGRGMDRGFGGRRGPGPG--PLMEOWGRRGRGPG--KMDXGE 646
Db 548 RCGAG-PPEGEGKGAAGPPGPAAGTGLQGMFERGGLSGFPGKDXGE 598

```

# RESULT 15

US-09-570-573-21

Sequence 21, Application US/09570573



Patent No. 6342361  
 GENERAL INFORMATION:  
 APPLICANT: Ovislc, Per  
 APPLICANT: Bonde, Martin  
 TITLE OF INVENTION: A Method for Assaying Collagen Fragments  
 TITLE OF INVENTION: in Body Fluids, A Test Kit and Means for Carrying Out the  
 TITLE OF INVENTION: Method and Use of the Method to Diagnose the Presence of  
 TITLE OF INVENTION: Disorders Associated with the Metabolism of  
 NUMBER OF SEQUENCES: 21  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Darby & Darby PC  
 STREET: 805 Third Avenue  
 CITY: New York  
 STATE: New York  
 COUNTRY: USA  
 ZIP: 10022  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patent in Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/09/570,573  
 FILING DATE:  
 CLASSIFICATION:  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: 08/187,319  
 FILING DATE:  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Gogorls, Adda C  
 REGISTRATION NUMBER: 29,714  
 REFERENCE/DOCKET NUMBER: 4305/08701  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 212-527-7700  
 TELEFAX: 212-753-6237  
 TELEX: 236687  
 INFORMATION FOR SRO ID NO: 21:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1078 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 ORIGINAL SOURCE:  
 ORGANISM: Homo sapiens  
 IMMEDIATE SOURCE:  
 CLONE: COLLAGEN ALPHA 1 (III)  
 IS-09-570-573-21

Query Match	10.4%	Score	376.5	DB	3	Length	1078
Best Local Similarity	25.8%	Pred. No.	8e-18				
Matches	184	Conservative	34	Mismatches	235	Indels	259
						Gaps	34
QY	42	GTGCGPTDVSYTOACTTATGCGTATATGCGPTGTTTPTAPOAQSOPVCGTGAVDTT	101				
DB	39	GTSHHP-----GSPGSPGVQSPGPR--GQAGSPSPG---PPAGT-----	75				
QY	102	TATVTTTTOASVTAASAVGTGPAPVAYGGQAPATAATPAPDGNKPTETSOPOSTGGYNOP	161				
DB	76	-----PSGPA-GDGESGRGRPDNDLP---GPPKIKPAPAP	110				
QY	162	SL-----GYGGSNYSTPQ-----VPSGY---PMQPTAAPSTPTPTYSSTQ	199				
DB	111	GFPCKMGKHGRDGRNGEKGTGAPGLKBNGLPGENGAAPGMPGRGAP-----	158				
QY	200	PTSYDQSSYSPONTNYGQPSSTYGQSSSYGQSSYGGQPTSYSPQSGYSQAQPSQSS	259				
DB	159	-----GERGRPLPGAAAGARNDARSGDQPPGP--PPGTAAGPSPGSG---AKG	204				
QY	260	SYGQSSSPRODHPSSMGVYGO--ESG--GFSG-----PGENRSMSPDNRG-----	301				
DB	205	EVGFAGS-----PSGNGAAPQGRGEGRPGGHAGAACQPPPPGLTNGSPGKKGEMPGACITPCA	259				
QY	302	-----RGRGFGDQGMNRGRGGRGGGAGMSAGERGFNKPCCGPMDEGDL	346				

Db 260 PGLMGARCP PGCPAGANGAPGL -RGAGGEPGKNGAGBPGRGERGAGIPGVPAAGAGEDG 318  
Qy 347 DLGPPVDEDEDDNSAIVVQGLNDSVTLDDLADLFFKQCGVVMKMKETQPMIHIYLDKET 406  
Db 319 KDGSQGDPCGANGLPALAGERALSS----- 343  
Qy 407 GKPKGDATVSYEDPPTAKAAVEWFDGDKFOGSKLVSLARKKP MNNSMG -GLPPRGR 464  
Db 344 RGPAGPNCIPBKGPAGE----- RGA RGPAGPRGAGAGEPGRDGV 382  
Qy 465 GMPPPLRGPGPGPGPG----- GPMGRNG--GRGDRCGFPPRGPGRSGT--- NPSG 512  
Db 383 PGCPGRMRMPSPGPGSGDGRKPGPGSGSQRPRPPG---PSGPGQGVWGFPGPKGN 439  
Qy 513 GNVHRADDMQCCPNNGCCNONPAMVTECNQCAKPRBEGLPPPPPP-----PGDRRRGR 568  
Db 440 DGAPBPKNGERGGP -GPPPG----- PPRKNGEYVQGPGRPGDGDKDDTGP 487  
Qy 569 GGMG-----GRGGLMDR-----GGPGMFRGGRGGDRG-----GP 599  
Db 488 RGPQQLQCLPQTGPGPGENBKPGEPKGEAGAPAPAGGKGDAAPGERGPGQLAGAPGL 547  
Qy 600 RGRGRMDRCGFGGGRGCGPGPG-----PLMEONGARRGRGRGPG--KMDKG 646  
Db 548 RGGAG-PPGPRGCGKGAAGPPGPFGAAGTFFGLQGMGERGGLGSPGPKDKGE 598

Search completed: February 18, 2005, 15:13:48  
Job time : 50 secs

**This Page Blank (uspto)**

---

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Comugen Ltd.

## OM protein - protein search, using sw model

Run on: February 18, 2005, 15:12:20 ; Search time 135 Seconds  
(without alignments)  
1590.149 Million cell updates/sec

Title: US-10-791-017A-2

Sequence: 1 MASTDSTYSQAAGGYSYA.....GGFGKDKGHHRRDRPY 656

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1380268 seqs, 327241040 residues

Total number of hits satisfying chosen parameters: 1380268

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

Published Applications AA:\*  
1: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep:\*  
2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep:\*  
3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep:\*  
4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep:\*  
5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep:\*  
6: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep:\*  
7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep:\*  
8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep:\*  
9: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pep:\*  
10: /cgn2\_6/ptodata/2/pubpaa/US09C\_PUBCOMB.pep:\*  
11: /cgn2\_6/ptodata/2/pubpaa/US09C\_NEW\_PUB.pep:\*  
12: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep:\*  
13: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB.pep:\*  
14: /cgn2\_6/ptodata/2/pubpaa/US10C\_PUBCOMB.pep:\*  
15: /cgn2\_6/ptodata/2/pubpaa/US10C\_PUBCOMB.pep:\*  
16: /cgn2\_6/ptodata/2/pubpaa/US10C\_PUBCOMB.pep:\*  
17: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep:\*  
18: /cgn2\_6/ptodata/2/pubpaa/US11\_NEW\_PUB.pep:\*  
19: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep:\*  
20: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3633	100.0	656	US-10-791-017A-2	Sequence 2, Appli
2	3277	90.2	600	US-10-094-749-2713	Sequence 2713, Ap
3	1281.5	35.3	525	US-10-755-889-650	Sequence 650, App
4	799.5	22.0	260	US-10-408-765A-116	Sequence 116, App
5	623	17.1	156	US-09-925-301-1511	Sequence 1511, Ap
6	475	13.1	83	US-09-864-761-42397	Sequence 42397, A
7	470.5	13.0	462	US-09-919-039-3324	Sequence 324, App
8	423	11.6	467	US-10-437-963-191926	Sequence 191926,
9	420	11.6	395	US-10-425-114-69744	Sequence 69744, A
10	400.5	11.0	1466	US-09-918-715-226	Sequence 226, App
11	400.5	11.0	1466	US-10-177-293-68	Sequence 68, Appl
12	400.5	11.0	1466	US-10-301-822-33	Sequence 33, Appl
13	400.5	11.0	1466	US-10-257-021-72	Sequence 72, Appl

14	400.5	11.0	1466	US-10-357-851-3	Sequence 3, Appli
15	400.5	11.0	1466	US-10-358-024-3	Sequence 3, Appli
16	400.5	11.0	1466	US-10-734-564-103	Sequence 103, App
17	388.5	10.7	1466	US-10-402-089-4	Sequence 4, Appli
18	388.5	10.7	1466	US-10-402-089-6	Sequence 6, Appli
19	388.5	10.7	1466	US-10-402-072A-4	Sequence 4, Appli
20	388.5	10.7	1466	US-10-402-072A-6	Sequence 6, Appli
21	384	10.6	1466	US-10-402-089-12	Sequence 12, Appli
22	384	10.6	1466	US-10-402-072A-12	Sequence 12, Appli
23	376.5	10.4	1078	US-10-058-124-21	Sequence 21, Appli
24	348.5	9.6	1449	US-10-402-089-8	Sequence 8, Appli
25	348.5	9.6	1449	US-10-402-072A-8	Sequence 8, Appli
26	348	9.6	1496	US-10-177-293-70	Sequence 70, Appli
27	348	9.6	1496	US-10-301-822-35	Sequence 35, Appli
28	348	9.6	1496	US-10-236-031B-74	Sequence 74, Appli
29	348	9.6	1496	US-10-468-091-22	Sequence 22, Appli
30	348	9.6	1496	US-10-788-792-248	Sequence 248, App
31	346.5	9.5	1496	US-10-437-963-164510	Sequence 164510,
32	346.5	9.5	1497	US-10-468-091-23	Sequence 23, Appli
33	346	9.5	1463	US-10-402-089-2	Sequence 2, Appli
34	346	9.5	1463	US-10-402-072A-2	Sequence 2, Appli
35	344	9.5	1453	US-10-468-091-26	Sequence 26, Appli
36	337.5	9.3	1461	US-10-468-091-25	Sequence 25, Appli
37	335.5	9.2	1418	US-10-058-124-20	Sequence 20, Appli
38	335.5	9.2	1418	US-10-468-091-5	Sequence 5, Appli
39	334	9.2	1739	US-09-795-061-2	Sequence 2, Appli
40	333.5	9.2	714	US-09-861-597-10	Sequence 10, Appli
41	333.5	9.2	714	US-10-887-100-10	Sequence 47, Appli
42	333	9.2	854	US-10-488-056-47	Sequence 261, App
43	333	9.2	1464	US-09-918-715-261	Sequence 159, App
44	333	9.2	1464	US-10-060-036-159	Sequence 36, Appli
45	333	9.2	1464	US-10-171-311-36	

## ALIGNMENTS

RESULT 1  
US-10-791-017A-2  
; Sequence 2, Application US/10791017A  
; Publication No. US20040197827A1  
; GENERAL INFORMATION:  
; APPLICANT: JENAPARM GmbH & Co. KG  
; TITLE OF INVENTION: Methods for Determining Hormonal Effects of Substances  
; FILE REFERENCE: Pat 3684/11  
; CURRENT APPLICATION NUMBER: US/10/791, 017A  
; CURRENT FILING DATE: 2004-03-02  
; NUMBER OF SEQ ID NOS: 7  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 2  
; LENGTH: 656  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-791-017A-2

Query Match 100.0%; Score 3633; DB 16; Length 656;  
Best Local Similarity 100.0%; Pred. No. 4.7e-209;  
Matches 656; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MASTDSTYSQAAGGYSYA...TPTGTAQTQAYGQOSYGYGPTDVSYQAQTTAT 60  
DB 1 MASTDSTYSQAAGGYSYA...TPTGTAQTQAYGQOSYGYGPTDVSYQAQTTAT 60  
QY YGOTATATSYGAPPTGTTPTTAPOASOPVQVGTGAYDTTATTTTQASYAQAAGAYGT 120  
DB YGOTATATSYGAPPTGTTPTTAPOASOPVQVGTGAYDTTATTTTQASYAQAAGAYGT 120  
QY 61 YGOTATATSYGAPPTGTTPTTAPOASOPVQVGTGAYDTTATTTTQASYAQAAGAYGT 120  
DB 61 YGOTATATSYGAPPTGTTPTTAPOASOPVQVGTGAYDTTATTTTQASYAQAAGAYGT 120  
QY 121 QPAPYVGGQPAATPAPTRPDGKPKETISQPOSSTGYNQPSIGYQSNYSYQVPGSY 180  
DB 121 QPAPYVGGQPAATPAPTRPDGKPKETISQPOSSTGYNQPSIGYQSNYSYQVPGSY 180  
QY 181 MGPVTPAPYPPPTPSYSTPSTDSYSGQNNYVGGQSSYGGQSSYVGGQPPPTSY 240  
DB 181 MGPVTPAPYPPPTPSYSTPSTDSYSGQNNYVGGQSSYGGQSSYVGGQPPPTSY 240

```

Db 181 MGPVTAPEPSYPTSTSTQPTSYDOSYSQONTYTGPPSSYGGQSSYGGQSSYGGQSPPTSY 240
Qy 241 PPTGTSYGAAPSOYSQOQSSSYGQSSFRQDHPSSMGVYGGESGFGSPGPNRSMSPDNR 300
Db 241 PPTGTSYGAAPSOYSQOQSSSYGQSSFRQDHPSSMGVYGGESGFGSPGPNRSMSPDNR 300
Qy 301 GRGRGDFDRGMSRGGRGGRGMSAGRGGRGPNKPGGPMDEGPDLDLGPVDPDESDN 360
Db 301 GRGRGDFDRGMSRGGRGGRGMSAGRGGRGPNKPGGPMDEGPDLDLGPVDPDESDN 360
Qy 361 SAIVYGLNDSTLDDIADFFKQCGVVKNKRTGQPMIHIYDKETGPKGDAVSYEDP 420
Db 361 SAIVYGLNDSTLDDIADFFKQCGVVKNKRTGQPMIHIYDKETGPKGDAVSYEDP 420
Qy 421 PTKAAVEMFDKDFGSKLXVSLARKKPPMNSMRGLPFRGRGMPPLRGPGGPGGR 480
Db 421 PTKAAVEMFDKDFGSKLXVSLARKKPPMNSMRGLPFRGRGMPPLRGPGGPGGR 480
Qy 481 GGPMSRGGRGGRGGRGPPRGSRGNPSGGGVNQHRAQDMQCPNPGCGNPFARTEC 540
Db 481 GGPMSRGGRGGRGGRGPPRGSRGNPSGGGVNQHRAQDMQCPNPGCGNPFARTEC 540
Qy 541 NCKAPKPEBGLPPLPPPGGDRGGRGPGGMRGGRGLMDRGPGGMFRGGRGDRGGR 600
Db 541 NCKAPKPEBGLPPLPPPGGDRGGRGPGGMRGGRGLMDRGPGGMFRGGRGDRGGR 600
Qy 601 GGRGMDRGGRGGRGGRGPPPLMEQMGRRGGRGGRGGRGGRGGRGGRGGRGGR 656
Db 601 GGRGMDRGGRGGRGGRGPPPLMEQMGRRGGRGGRGGRGGRGGRGGRGGRGGR 656

```

## RESULT 2

```

US-10-094-749-2713
; Sequence 2713, Application US/10094749
; Publication No. US20030219741A1
; GENERAL INFORMATION:
; APPLICANT: ISOGAI, TAKAO
; APPLICANT: SUGIYAMA, TOMOYASU
; APPLICANT: OTSUKI, TETSUJI
; APPLICANT: WAKAMATSU, AI
; APPLICANT: SATO, HIROYUKI
; APPLICANT: ISHII, SHIZUKO
; APPLICANT: YAMAMOTO, JUN-ICHI
; APPLICANT: ISONO, YUUKO
; APPLICANT: 'HIO, YURI
; APPLICANT: OTSUKA, KAORU
; APPLICANT: NAGAI, KEIICHI
; APPLICANT: IRIE, RYOTARO
; APPLICANT: TAMECHIKA, ICHIRO
; APPLICANT: SEKI, NAOHITO
; APPLICANT: YOSHIKAWA, TSUTOMU
; APPLICANT: OTSUKA, MOTOTYUKI
; APPLICANT: NAGAHARI, KENJI
; APPLICANT: MASUHO, YASUHIKO
; TITLE OF INVENTION: NOVEL FULL-LENGTH CDNA
; FILE REFERENCE: 084335/0160
; CURRENT APPLICATION NUMBER: US/10/094,749
; CURRENT FILING DATE: 2002-03-12
; PRIOR APPLICATION NUMBER: 60/350,435
; PRIOR FILING DATE: 2002-01-24
; PRIOR APPLICATION NUMBER: JP 2001-328381
; PRIOR FILING DATE: 2001-09-14
; NUMBER OF SEQ ID NOS: 3381
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2713
; LENGTH: 600
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-094-749-2713

```

```

Query Match 90.2%; Score 3277; DB 15; Length 600;
Best Local Similarity 91.3%; Pred. No. 7.9e-188;
Matches 599; Conservative 0; Mismatches 1; Indels 56; Gaps 1;

```

```

Qy 1 MASTDSTYSQAQAAGQVSAVTAAPPTGYAQTTOAIVGQSSYGVYGGQPTDVSYTAQTTAT 60
Db 1 MASTDSTYSQAQAAGQVSAVTAAPPTGYAQTTOAIVGQSSYGVYGGQPTDVSYTAQTTAT 60
Qy 61 YGOTAVATSYGQPTGYTTPTTAPOAVSQPVQYGTGAVDTTATVTTTAAVAAQSAVGT 120
Db 61 YGOTAVATSYGQPTGYTTPTTAPOAVSQPVQYGTGAVDTTATVTTTAAVAAQSAVGT 120
Qy 121 GPAYPAYGQOPATATPATRQDQGNKPTETSQPOSSTGTVNQPSLGVQSNYSYIPQVGSY 180
Db 121 GPAYPAYGQOPATATPATRQDQGNKPTETSQPOSSTGTVNQPSLGVQSNYSYIPQVGSY 180
Qy 181 MGPVTAPEPSYPTSTSTQPTSYDOSYSQONTYTGPPSSYGGQSSYGGQSSYGGQSPPTSY 240
Db 181 MGPVTAPEPSYPTSTSTQPTSYDOSYSQONTYTGPPSSYGGQSSYGGQSSYGGQSPPTSY 240
Qy 241 PPTGTSYGAAPSOYSQOQSSSYGQSSFRQDHPSSMGVYGGESGFGSPGPNRSMSPDNR 300
Db 241 PPTGTSYGAAPSOYSQOQSSSYGQSSFRQDHPSSMGVYGGESGFGSPGPNRSMSPDNR 300
Qy 301 GRGRGDFDRGMSRGGRGGRGMSAGRGGRGPNKPGGPMDEGPDLDLGPVDPDESDN 360
Db 301 GRGRGDFDRGMSRGGRGGRGMSAGRGGRGPNKPGGPMDEGPDLDLGPVDPDESDN 360
Qy 361 SAIVYGLNDSTLDDIADFFKQCGVVKNKRTGQPMIHIYDKETGPKGDAVSYEDP 420
Db 361 SAIVYGLNDSTLDDIADFFKQCGVVKNKRTGQPMIHIYDKETGPKGDAVSYEDP 420
Qy 421 PTKAAVEMFDKDFGSKLXVSLARKKPPMNSMRGLPFRGRGMPPLRGPGGPGGR 480
Db 421 PTKAAVEMFDKDFGSKLXVSLARKKPPMNSMRGLPFRGRGMPPLRGPGGPGGR 480
Qy 481 GGPMSRGGRGGRGGRGPPRGSRGNPSGGGVNQHRAQDMQCPNPGCGNPFARTEC 540
Db 481 GGPMSRGGRGGRGGRGPPRGSRGNPSGGGVNQHRAQDMQCPNPGCGNPFARTEC 540
Qy 541 NCKAPKPEBGLPPLPPPGGDRGGRGPGGMRGGRGLMDRGPGGMFRGGRGDRGGR 600
Db 541 NCKAPKPEBGLPPLPPPGGDRGGRGPGGMRGGRGLMDRGPGGMFRGGRGDRGGR 600
Qy 601 GGRGMDRGGRGGRGGRGPPPLMEQMGRRGGRGGRGGRGGRGGRGGRGGRGGR 656
Db 601 GGRGMDRGGRGGRGGRGPPPLMEQMGRRGGRGGRGGRGGRGGRGGRGGRGGR 656

```

## RESULT 3

```

US-10-755-889-650
; Sequence 650, Application US/10755889
; Publication No. US20040171823A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES ASSOCIATED WITH THE NF-KB
; FILE REFERENCE: D0284 NP
; CURRENT APPLICATION NUMBER: US/10/755,889
; CURRENT FILING DATE: 2004-01-13
; PRIOR APPLICATION NUMBER: U.S. 60/440,068
; PRIOR FILING DATE: 2003-01-14
; PRIOR APPLICATION NUMBER: U.S. 60/469,757
; PRIOR FILING DATE: 2003-05-12
; NUMBER OF SEQ ID NOS: 823
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 650
; LENGTH: 525
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-755-889-650

```

```

Query Match 35.3%; Score 1281.5; DB 16; Length 525;
Best Local Similarity 44.8%; Pred. No. 1.1e-68;
Matches 299; Conservative 59; Mismatches 155; Indels 155; Gaps 29;

```

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

# OM protein - protein search, using SW model

Run on: February 18, 2005, 14:57:35 ; Search time 44 Seconds  
(Without alignments)  
1434.504 Million cell updates/sec

Title: US-10-791-017a-2

Perfect score: 3633  
Sequence: 1 MASTDSTYSQAAGGYSYA.....GGGKMDKGEHNGERDRPY 656

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

PIR 79: \*  
1: PIR: \*  
2: PIR: \*  
3: PIR: \*  
4: PIR: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3633	100.0	656	1 A49358	RNA-binding protei
2	3571.5	98.3	655	1 A55726	RNA-binding protei
3	1282	35.3	526	1 S33799	RNA-binding protei
4	1270.5	35.0	528	2 G02127	fus-like protein -
5	1003.5	27.6	589	2 S71954	RNA/sgDNA-binding
6	724	19.9	404	2 S54729	RNA-binding protei
7	554	15.2	545	2 T1567	hypothetical prote
8	484	13.3	295	4 S33798	RNA binding protei
9	472.5	13.0	462	4 S33798	FUS/CHOP mutant fu
10	400.5	11.0	1466	1 CGH7L	collagen alpha 1(I
11	376.5	10.4	886	2 T50694	collagen alpha 1(I
12	374	10.3	1464	2 S59856	collagen alpha 1(I
13	370.5	10.2	211	2 C96539	hypothetical prote
14	368	10.1	1049	1 CGB07S	collagen alpha 1(I
15	359	9.9	1838	1 CGH01V	collagen alpha 1(I
16	348	9.6	1496	1 CGH02V	collagen alpha 2(V
17	346.5	9.5	1497	2 T49607	procollagen type V
18	344	9.5	1453	2 S21626	collagen alpha 1(I
19	344	9.5	1843	2 S18803	collagen alpha 1(V
20	343.5	9.5	1880	2 T18531	traccin - medicina
21	342.5	9.4	1414	1 S23809	collagen alpha 2(I
22	340.5	9.4	848	2 S02262	glutinin high mole
23	339.5	9.3	1034	2 S18251	collagen alpha 1(X
24	339.5	9.3	1482	2 A40333	collagen alpha 1(I
25	338.5	9.3	1486	1 B40333	collagen alpha 1(I
26	336.5	9.3	1418	2 T45467	collagen alpha 1(I
27	336	9.2	838	1 EEWTHM	glutinin, high mol
28	336	9.2	1752	2 A45407	collagen alpha 3(I
29	335.5	9.2	1487	1 CGH06C	collagen alpha 1(I

30	333	9.2	1464	1 CGH01S	collagen alpha 1(I
31	330	9.1	1806	1 CGH01E	collagen alpha 1(X
32	328	9.0	1027	2 S28774	collagen alpha cha
33	328	9.0	2944	2 A54849	collagen alpha 1(V
34	324	8.9	671	1 CGR1S	collagen alpha 1(I
35	324	8.9	1585	2 T31611	hypothetical prote
36	322.5	8.9	1747	2 A54121	collagen alpha-4 c
37	321.5	8.8	3198	2 A43426	collagen alpha 2 f
38	320.5	8.8	1042	1 CGCH1S	collagen alpha 1(I
39	318	8.8	1142	2 T00022	B120 protein - hum
40	317	8.7	1419	2 A41182	collagen alpha 1(I
41	317	8.7	1487	2 B41182	collagen alpha 1(I
42	316.5	8.7	1670	1 CGH13B	collagen alpha 3(I
43	316	8.7	753	2 JC2099	glutinin, high mol
44	316	8.7	921	2 S40495	collagen alpha 1(I
45	315	8.7	921	2 S42617	collagen alpha 1(I

## ALIGNMENTS

RESULT 1  
A49358  
RNA-binding protein EMS - human  
C/Species: Homo sapiens (man)  
C/Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 09-Jul-2004  
C/Accession: A49358; S28257  
R/Plougaestel, B.; Zucman, J.; Peter, M.; Thomas, G.; Delattre, O.  
Genomics 18, 609-615, 1993  
A/Title: Genomic structure of the EMS gene and its relationship to EMSR1, a site of tumor  
A/Reference number: A49358; MIM:9410360; PMID:8307570  
A/Accession: A49358  
A/Status: nucleic acid sequence not shown; translation not shown  
A/Molecule type: DNA  
A/Residues: 1-656 <RES>  
A/Cross-references: UNIPROT:001844; EMBL:X72990; NID:9485838; PIR:CAA51489.1; PID:982565  
R/Delattre, O.; Zucman, J.; Plougaestel, B.; Desmaziere, C.; Melot, T.; Peter, M.; Kovar, H.,  
Nature 359, 162-165, 1992  
A/Title: Gene fusion with an ETS DNA-binding domain caused by chromosome translocation in  
A/Reference number: S28257; MIM:92396239; PMID:1522903  
A/Accession: S28257  
A/Molecule type: mRNA  
A/Residues: 1-656 <DEL>  
A/Cross-references: EMBL:X66899; NID:9547565; PIR:CAA47350.1; PID:931280  
C/Genetics:  
A/Genes: GDB:EMSR1  
A/Map position: 22q12.1-22q12.1  
A/Introns: 5/1; 17/2; 34/3; 76/1; 138/2; 265/1; 325/2; 338/1; 349/1; 388/3; 432/1.  
A/Note: EMSR1 region is exons 7-10 of this gene, called EMS in reference A49358; this reg  
C/Keywords: carcinogenesis; nucleus; RNA binding; tandem repeat  
F/662-437/Domain: ribonucleoprotein repeat homology <RMS>

Query Match 100.0%; Score 3633; DB 1; Length 656;  
Best Local Similarity 100.0%; Pred. No. 3.9e-166;  
Matches 656; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MASTDSTYSQAAGGYSYAQTQAQAGGOSYGYGPTDVSYQAQTTAT	60
DB	1	MASTDSTYSQAAGGYSYAQTQAQAGGOSYGYGPTDVSYQAQTTAT	60
QY	61	YGGTAATATSYGQPTGTTTPTAPQAVSOPVQVGTGAVDTTATVTTTQASVAAGSAYGT	120
DB	61	YGGTAATATSYGQPTGTTTPTAPQAVSOPVQVGTGAVDTTATVTTTQASVAAGSAYGT	120
QY	121	QPAVPAVGGQPAATATATRRQDGNKPTETGQPOSSTGYNQPSLGGQSNYSYQVVGSGYP	180
DB	121	QPAVPAVGGQPAATATATRRQDGNKPTETGQPOSSTGYNQPSLGGQSNYSYQVVGSGYP	180
QY	181	MCPTAPSPVPTPSYSTPTSDSSYSQONTYGGQSSYGGQSSYGGQSSYGGQSSYGGQSSY	240
DB	181	MCPTAPSPVPTPSYSTPTSDSSYSQONTYGGQSSYGGQSSYGGQSSYGGQSSYGGQSSY	240

QY 241 PPTGTSYSAAPSOYSSQSSSSYGGSSSFRDHPSSMGVYQGESGSGFSGPGENRSMSPDNR 300  
 DB 241 PPTGTSYSAAPSOYSSQSSSSYGGSSSFRDHPSSMGVYQGESGSGFSGPGENRSMSPDNR 300  
 QY 301 GRGGGFDGRGMSRGGRGGRGMSAGRGFNKPGPMDEBPDLDLGPVDPDESDN 360  
 DB 301 GRGGGFDGRGMSRGGRGGRGMSAGRGFNKPGPMDEBPDLDLGPVDPDESDN 360  
 QY 361 SATYVGGLNDSTLTDLDLADFFKQCGVYKNNKRTGQPMHIIYLDKKGKPKGDAVSYEDP 420  
 DB 361 SATYVGGLNDSTLTDLDLADFFKQCGVYKNNKRTGQPMHIIYLDKKGKPKGDAVSYEDP 420  
 QY 421 PTKAAVWFVFDGDFGSSKLVSLARKKPPMNSMRGGLPRERGRMPPLRGPGPGGP 480  
 DB 421 PTKAAVWFVFDGDFGSSKLVSLARKKPPMNSMRGGLPRERGRMPPLRGPGPGGP 480  
 QY 481 GCPMGMRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGR 540  
 DB 481 GCPMGMRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGR 540  
 QY 541 NOCKAPKPEBGLPPLPPPPPGDRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGR 600  
 DB 541 NOCKAPKPEBGLPPLPPPPPGDRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGR 600  
 QY 601 GGRGMDRGFGGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGR 656  
 DB 601 GGRGMDRGFGGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGR 656

## RESULT 2

A55726  
 RNA-binding protein Ews - mouse  
 C/Species: Mus musculus (house mouse)  
 C/Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 09-Jul-2004  
 C/Accession: A55726; S45007  
 R/Plouffe, B.; Mattei, M.G.; Thomas, G.; Delattre, O.  
 Genomics 23, 278-281, 1994  
 A/Title: Cloning and chromosome localization of the mouse Ews gene.  
 A/Reference number: A55726; MUID:95130099; PMID:7829090  
 A/Accession: A55726  
 A/Status: preliminary  
 A/Molecule type: mRNA  
 A/Residues: 1-655 <PRO>  
 A/Cross-references: UNIPROT:Q61545; GB:X79233; NID:g488512; PIND:CA55815.1; PID:g488513  
 A/Note: authors translated the codon TCA for residue 116 as Thr, GCC for residue 123 as C;Genetic:  
 A:Gene: Ews  
 C/Superfamily: RNA-binding protein, EWS type; ribonucleoprotein repeat homology  
 C/Keywords: carcinogenesis; nucleus; RNA binding; tandem repeat  
 F/361-436/Domain: ribonucleoprotein repeat homology <RMB>

Query Match 98.3%; Score 3571.5; DB 1; Length 655;  
 Best Local Similarity 98.0%; Pred. No. 3.6e-165;  
 Matches 643; Conservative 8; Mismatches 4; Indels 1; Gaps 1;  
 QY 1 MASTDSTVSQAAGQSYAATPOTGVAOTTAAGQSSGTYGQPTDVSYTQAOTTA 60  
 DB 1 MASTDSTVSQAAGQSYAATPOTGVAOTTAAGQSSGTYGQPTDVSYTQAOTTA 60  
 QY 61 YGOTATATSYGQPTGTTTPTAFQAYSQPVQYGTGAYDTTATVTTTQASVAAQAYGT 120  
 DB 61 YGOTATATSYGQPTGTTTPTAFQAYSQPVQYGTGAYDTTATVTTTQASVAAQAYGT 120  
 QY 121 QPAYPAVGGQPAATPAPTRPDGKPTETSGPSSTGGYVNPGLGYGQSNYSYQVPGSY 180  
 DB 121 QPAYPAVGGQPAATPAPTRPDGKPTETSGPSSTGGYVNPGLGYGQSNYSYQVPGSY 180  
 QY 181 MGVNTPAPSPPTSYST 240  
 DB 181 MGVNTPAPSPPTSYST 240  
 QY 241 PPTGTSYSAAPSOYSSQSSSSYGGSSSFRDHPSSMGVYQGESGSGFSGPGENRSMSPDNR 300

DB 241 PPTGTSYSAAPSOYSSQSSSSYGGSSSFRDHPSSMGVYQGESGSGFSGPGENRSMSPDNR 300  
 QY 301 GRGGGFDGRGMSRGGRGGRGMSAGRGFNKPGPMDEBPDLDLGPVDPDESDN 360  
 DB 301 GRGGGFDGRGMSRGGRGGRGMSAGRGFNKPGPMDEBPDLDLGPVDPDESDN 360  
 QY 361 SATYVGGLNDSTLTDLDLADFFKQCGVYKNNKRTGQPMHIIYLDKKGKPKGDAVSYEDP 420  
 DB 361 SATYVGGLNDSTLTDLDLADFFKQCGVYKNNKRTGQPMHIIYLDKKGKPKGDAVSYEDP 420  
 QY 421 PTKAAVWFVFDGDFGSSKLVSLARKKPPMNSMRGGLPRERGRMPPLRGPGPGGP 480  
 DB 421 PTKAAVWFVFDGDFGSSKLVSLARKKPPMNSMRGGLPRERGRMPPLRGPGPGGP 480  
 QY 481 GCPMGMRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGR 540  
 DB 481 GCPMGMRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGR 540  
 QY 541 NOCKAPKPEBGLPPLPPPPPGDRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGR 600  
 DB 541 NOCKAPKPEBGLPPLPPPPPGDRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGR 600  
 QY 601 GGRGMDRGFGGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGR 656  
 DB 601 GGRGMDRGFGGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGR 656

## RESULT 3

S33799  
 RNA-binding protein FUS, nuclear - human  
 N/Alternate names: RNA-binding protein TLS  
 C/Species: Homo sapiens (man)  
 C/Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 09-Jul-2004  
 C/Accession: S33799; S36157  
 R/Crozat, A.; Aman, P.; Mandahl, N.; Ron, D.  
 Nature 363, 640-644, 1993  
 A/Title: Fusion of CHOP to a novel RNA-binding protein in human myxoid liposarcoma.  
 A/Reference number: S33799; MUID:93288139; PMID:8510758  
 A/Accession: S33799  
 A/Molecule type: mRNA  
 A/Residues: 1-526 <CRO>  
 A/Cross-references: UNIPROT:P35637; GB:S62140; NID:g386156; PIND:AAB27102.1; PID:g386157  
 A/Experimental source: liposarcoma  
 R/Rabbits, T.H.; Porter, A.; Larson, R.; Nathan, P.  
 Nature Genet. 4, 175-180, 1993  
 A/Title: Fusion of the dominant negative transcription regulator CHOP with a novel gene  
 A/Reference number: S36157; MUID:93350637; PMID:7503811  
 A/Accession: S36157  
 A/Status: preliminary  
 A/Molecule type: mRNA  
 A/Residues: 1-63, 'S', 66-526 <RAB>  
 A/Cross-references: EMBL:X71428; NID:g393415; PIND:CA50559.1; PID:g4210363  
 A/Experimental source: liposarcoma  
 C/Genetic:  
 A:Gene: GDB:FUS  
 A/Cross-references: GDB:136048; OMIM:137070  
 A/Map position: 16p11.2-16p11.2  
 C/Function:  
 A/Description: RNA-binding; probable plays a role in transcriptional regulation  
 C/Superfamily: RNA-binding protein, EWS type; ribonucleoprotein repeat homology  
 C/Keywords: carcinogenesis; nucleus; RNA binding; tandem repeat  
 F/286-361/Domain: ribonucleoprotein repeat homology <RMB>

Query Match 35.3%; Score 1282; DB 1; Length 526;  
 Best Local Similarity 45.0%; Pred. No. 4.7e-55;  
 Matches 301; Conservative 59; Mismatches 153; Indels 156; Gaps 30;  
 QY 1 MASTDSTVSQAAGQSYAATPOTGVAOTTAAGQSSGTYGQPTDVSYTQAOTTA 59  
 DB 1 MASTDSTVSQAAGQSYAATPOTGVAOTTAAGQSSGTYGQPTDVSYTQAOTTA 59  
 QY 60 TVGOTATATSYGQPTGTTTPTAFQAYSQPVQYGTGAYDTTATVTTTQASVAAQAYGT 117



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 18, 2005, 14:55:56 ; Search time 177 Seconds

(without alignments)  
1897.875 Million cell updates/sec

Title: US-10-791-017A-2

Perfect score: 3633

Sequence: 1 MASTDYSTYSQAAAQGGTGA.....GGPRKMDKSHRQRRDRPY 656

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : UniProt 03:\*

1: uniprot\_sprot.\*

2: uniprot\_trembl.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysts of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3633	100.0	656	1	EMS_HUMAN
2	3618.5	99.6	655	2	Q96FE8
3	3605.5	99.2	661	2	Q96MX4
4	3579	98.5	656	2	Q6NVA3
5	3571.5	98.3	655	1	EMS_MOUSE
6	3564.5	98.1	655	2	Q9D3P0
7	3277	90.2	600	2	Q96MNA
8	2601	71.6	674	2	Q6F3N0
9	2586.5	71.2	673	2	Q6DJQ4
10	1885.5	51.9	624	2	Q6NWX1
11	1874	51.6	633	2	Q9CRS5
12	1865	51.3	623	2	Q7SZR6
13	1753	48.3	354	2	Q9BWA2
14	1579	43.1	578	2	Q803B3
15	1566	43.5	575	2	Q6NVA7
16	1439	39.3	476	2	Q9BZD1
17	1332	36.4	504	2	Q6J4Y8
18	1332	36.4	536	2	Q7ZXQ2
19	1301	35.8	512	1	FUS_BOVIN
20	1290	35.5	518	1	FUS_MOUSE
21	1290	35.5	536	2	Q6DKB4
22	1289.5	35.5	517	2	Q8CFQ9
23	1282	35.3	526	1	FUS_HUMAN
24	1277	35.2	526	2	Q8TBR3
25	1270.5	35.0	528	2	Q13344
26	1016.5	28.0	557	2	Q8BQ46
27	1009	27.8	532	1	RB56_HUMAN
28	910	25.0	337	2	Q95XG8
29	907	25.0	337	2	Q86X94
30	904.5	24.9	501	2	Q6GDD1
31	900.5	24.8	482	2	Q6DCR4

32	884.5	24.3	475	2	Q66J39
33	782	21.5	410	2	Q7ZUE3
34	717.5	19.7	399	1	CAZ_DROME
35	690	19.0	442	2	Q7PUK2
36	625.5	17.2	189	2	Q7IE78
37	621	17.1	280	2	Q91VQ2
38	583.5	16.1	448	2	Q18265
39	452.5	12.5	265	2	Q6DDA7
40	423	11.6	414	2	Q69TN3
41	420.5	11.6	101	2	Q98UE8
42	414	11.4	394	2	Q6YS82
43	400.5	11.0	1163	2	Q8N6U4
44	400.5	11.0	1466	1	CA13_HUMAN
45	400	11.0	113	2	Q8BQ4
					Q8BQ4 mus musculus

## ALIGNMENTS

RESULT 1  
EMS\_HUMAN STANDARD; PRT; 656 AA.  
AC 001844; Q92635;  
DT 01-JUN-1994 (Rel. 29, Last sequence update)  
DT 01-JUN-1994 (Rel. 29, Last sequence update)  
DT 25-OCT-2004 (Rel. 45, Last annotation update)  
DE RNA-binding protein EMS (EWS oncogene) (Ewing sarcoma breakpoint region 1 protein).  
GN Name=EWSR1; Synonyms=EWS;  
OS Homo sapiens (human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Petal brain;  
RX MEDLINE=92396239; PubMed=1522903; DOI=10.1038/359162a0;  
RA Delattre O., Zucman J., Plougaerel B., Desmaziere C., Melot T., Peter M., Kovar H., Joubert I., de Jong P., Rouleau G., Aurias A., Thomas G.;  
RT "Gene fusion with an ETS DNA-binding domain caused by chromosome translocation in human tumours.";  
RT Nature 359:162-165(1992).  
RL  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=94140360; PubMed=8307570;  
RA Plougaerel B., Zucman J., Peter M., Thomas G., Delattre O.;  
RT "Genomic structure of the EWS gene and its relationship to EWSR1, a site of tumor-associated chromosome translocation.";  
RT Genome 18:609-615(1993).  
RL  
RN [3]  
RP SEQUENCE FROM N.A.  
RX Zucman-Rossi J.;  
RT Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.  
RL [4]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Placenta;  
RX MEDLINE=2288257; PubMed=12477932; DOI=10.1073/pnas.242603899;  
RA Krausberg R.L., Feingold E.A., Grouse L.H., Derge J.G., Stenhouse R.D., Collins F.S., Wagner L., Shennan C.M., Schuler G.D., Altshuler S.F., Zeeberg B., Buettow K.H., Schaefer C.F., Bhat N.K., Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L., Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E., Brownstein M.J., Ueda T.B., Tohyuki S., Carninci P., Prange C., Bosak S.S., Loughran N.A., Peters G.J., Abramson R.D., Mullany S.J., Rohak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H., Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W., Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A., Fahey J., Helton E., Kettlemen M., Madan A., Rodriguez S., Sanchez A., Whitting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G., Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C., Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalins D.E.,

RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length human  
 RT and mouse cDNA sequences.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 RN [5]  
 RX SEQUENCE OF 1-345 FROM N.A.  
 RX MEDLINE=97131501; PubMed=8975669; DOI=10.1006/geno.1996.0625;  
 RA Zucman-Rossi J., Legoux P., Thomas G.;  
 RT "Identification of new members of the Gas2 and Ras families in the  
 RT 22q12 chromosome region.";  
 RL Genomics 38:247-254(1996).  
 RN [6]  
 RX SEQUENCE OF 241-268 FROM N.A.  
 RX TISSUE=Placenta;  
 RX MEDLINE=95352541; PubMed=7542907;  
 RA Bhagirath T., Abe S., Nojima T., Yoshida M.C.;  
 RT "Molecular analysis of a t(11;22) translocation junction in a case of  
 RT Ewing's sarcoma.";  
 RL Genes Chromosomes Cancer 13:126-132(1995).  
 RN [7]  
 RX PHOSPHORYLATION SITE SER-266 AND PARTIAL SEQUENCE.  
 RX MEDLINE=98001723; PubMed=9341188; DOI=10.1074/jbc.272.43.27369;  
 RA Deloume J.C., Pritchard L., Delattre O., Storm D.R.;  
 RT "The proto-oncoprotein EWS binds calmodulin and is phosphorylated by  
 RT protein kinase C through an IQ domain.";  
 RL J. Biol. Chem. 272:27369-27377(1997).  
 RN [8]  
 RX SEQUENCE OF 128-158, 233-247, 268-324, 334-364, 393-439, 447-518 AND  
 RX 551-641, METHYLATION OF ARGININES, AND MASS SPECTROMETRY.  
 RX MEDLINE=1276345; PubMed=11278906; DOI=10.1074/jbc.M011446200;  
 RA Belyanskaya L.L., Gehrig P.M., Gehrig H.;  
 RT "Exposure on cell surface and extensive arginine methylation of Ewing  
 RT sarcoma (EWS) protein.";  
 RL J. Biol. Chem. 276:18661-18667(2001).  
 RN [9]  
 RX ALTERNATIVE SPLICING, AND RNA-BINDING.  
 RX MEDLINE=94366763; PubMed=8084618;  
 RA Ono T., Ouchida M., Lee L., Gatalica Z., Rao V.N., Reddy B.S.P.;  
 RT "The EWS gene, involved in Ewing family of tumors, malignant melanoma  
 RT of soft parts and desmoplastic small round cell tumors, codes for an  
 RT RNA binding protein with novel regulatory domains.";  
 RL Oncogene 9:3087-3097(1994).  
 RN [10]  
 RX INTERACTION WITH SP1.  
 RX MEDLINE=98355009; PubMed=9660765; DOI=10.1074/jbc.273.29.18086;  
 RA Zhang D., Paley A.J., Childs G.;  
 RT "The transcriptional repressor ZFM1 interacts with and modulates the  
 RT ability of EWS to activate transcription.";  
 RL J. Biol. Chem. 273:18086-18091(1998).  
 RN [11]  
 RX CHARACTERIZATION.  
 RX MEDLINE=20390060; PubMed=10767297; DOI=10.1074/jbc.M002961200;  
 RA Li K.K.C., Lee K.A.W.;  
 RT "Transcriptional activation by the Ewing's sarcoma (EWS) oncogene can  
 RT be cis-repressed by the EWS RNA-binding domain.";  
 RL J. Biol. Chem. 275:23053-23058(2000).  
 CC -1- FUNCTION: Might normally function as a repressor. EWS-fusion-  
 CC proteins (EFPs) may play a role in the tumorigenic process. They  
 CC may disturb gene expression by mimicking, or interfering with the  
 CC normal function of CTD-POLII within the transcription initiation  
 CC complex. They may also contribute to an aberrant activation of the  
 CC fusion protein target genes.  
 CC -1- SUBUNIT: Binds POLR2C, SP1, calmodulin and RNA. Interacts with  
 CC PRK2B/PAK2.  
 CC -1- SUBCELLULAR LOCATION: Nuclear, cytoplasmic and on cell surface.  
 CC Relocates from cytoplasm to ribosomes upon PRK2B/PAK2 activation.  
 CC -1- ALTERNATIVE PRODUCTS:  
 CC Event=Alternative splicing; Named isoforms=2;  
 CC Name=EWS;  
 CC IsoId=001844-1; Sequence=Displayed;  
 CC Name=EWS-B;  
 CC IsoId=001844-2; Sequence=VSP\_005793;  
 CC Note=No experimental confirmation available;

CC TISSUE SPECIFICITY: Ubiquitous.  
 CC -1- DOMAIN: EWS activation domain (EAD) functions as a potent  
 CC activation domain in EFPs. EWS1 binds POLR2C but not POLR2E or  
 CC POLR2G, whereas the isolated EAD binds POLR2E and POLR2G but not  
 CC POLR2C. Cis-linked RNA-binding domain (RBD) can strongly and  
 CC specifically repress trans-activation by the EAD.  
 CC -1- PTM: Phosphorylated; calmodulin-binding inhibits phosphorylation  
 CC of Ser-266.  
 CC -1- DISBASE: Ewing's sarcoma is characterized by chromosomal  
 CC translocations t(11;22)(q24;q12) which involves EWS1 and FLI1,  
 CC t(7;22)(p22;q12) which involves EWS1 and ETV1, t(21;22)(q22;q12)  
 CC which involves EWS1 and ERG and t(9;22)(q22-31;q1-12) which  
 CC involves EWS1 and NR4A3.  
 CC -1- DISBASE: Involved in desmoplastic small round cell tumor (DSRCT)  
 CC through a chromosomal translocation t(11;22)(p13;q12) that  
 CC involves EWS1 and WT1.  
 CC -1- DISBASE: Malignant melanoma of soft parts (MMP), also known as  
 CC soft tissue clear cell sarcoma, is a rare tumor developing in  
 CC tendons and aponeuroses. It is associated with chromosomal  
 CC translocation t(12;22)(q13;q12) involving EWS1 and ATF-1.  
 CC -1- DISBASE: Involved in small round cell sarcoma through a  
 CC chromosomal translocation t(11;22)(p36.1;q12) that involves EWS1  
 CC and ZNF78.  
 CC -1- MISCELLANEOUS: EFPs arise due to chromosomal translocations in  
 CC which EWS1 is fused to a variety of cellular transcription  
 CC factors. EFPs are very potent transcriptional activators dependent  
 CC on the EAD and a C-terminal DNA-binding domain contributed by the  
 CC fusion partner. The spectrum of malignancies associated with EFPs  
 CC are thought to arise via EFP-induced transcriptional deregulation,  
 CC with the tumor phenotype specified by the EWS1 fusion partner and  
 CC cell type. Transcriptional repression of the transforming growth  
 CC factor beta type II receptor (TGF beta RI) is an important target  
 CC of the EWS-FLI1, EWS-ERG, or EWS-ETV1 oncogene.  
 CC -1- MISCELLANEOUS: Binds calmodulin in the presence, but not in the  
 CC absence, of calcium ion.  
 CC -1- SIMILARITY: Belongs to the RNP TET family.  
 CC -1- SIMILARITY: Contains 1 IQ domain.  
 CC -1- SIMILARITY: Contains 1 RanBP2-type zinc finger.  
 CC -1- SIMILARITY: Contains 1 RNA recognition motif (RRM) domain.  
 CC -1- DATABASES: NMR=Atlas Genet. CytoGenet. Oncol. Haematol.;  
 CC WWW="http://www.infodolgen.fr/services/chromocancer/Genes/EWS11085.html".  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: X66899; CAA47350.1; -  
 CC EMBL: X72990; CAA51489.1; -  
 CC EMBL: X72991; CAA51489.1; JOINED.  
 CC EMBL: X72992; CAA51489.1; JOINED.  
 CC EMBL: X72993; CAA51489.1; JOINED.  
 CC EMBL: X72994; CAA51489.1; JOINED.  
 CC EMBL: X72995; CAA51489.1; JOINED.  
 CC EMBL: X72996; CAA51489.1; JOINED.  
 CC EMBL: X72997; CAA51489.1; JOINED.  
 CC EMBL: X72998; CAA51489.1; JOINED.  
 CC EMBL: X72999; CAA51489.1; JOINED.  
 CC EMBL: X73000; CAA51489.1; JOINED.  
 CC EMBL: X73001; CAA51489.1; JOINED.  
 CC EMBL: X73002; CAA51489.1; JOINED.  
 CC EMBL: X73003; CAA51489.1; JOINED.  
 CC EMBL: X73004; CAA51489.1; JOINED.  
 CC EMBL: BC004817; AAH04817.1; -  
 CC EMBL: Y08806; CAA70044.1; ALT\_INIT.  
 CC EMBL: AB016435; BAA31990.1; -  
 CC EMBL: Y07848; CAA69177.1; -  
 CC EMBL: Y07848; CAA69177.1; -  
 CC PIR: A49358; A49358.  
 CC HSSP: O95218; I102.  
 CC Genew; HGNC:3508; EWS1.

DR H-InvDB: HIT0016349; --  
DR MIM; 133450; --

Query Match 100.0%; Score 3633; DB 1; Length 656;  
Best Local Similarity 100.0%; Pred. No. 2, 7e-148;  
Matches 656; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MASTDSTYSQAQAAGGYSATYTAQPTGTAQTTQAVGQSSYGYGPTDVSYTAQTTAT 60  
DB 1 MASTDSTYSQAQAAGGYSATYTAQPTGTAQTTQAVGQSSYGYGPTDVSYTAQTTAT 60  
QY 61 YGQTAATATSGQPTGTTPTAQAASQPVQSGTGAVDTTTATVTTTQASVAQAAGYGT 120  
DB 61 YGQTAATATSGQPTGTTPTAQAASQPVQSGTGAVDTTTATVTTTQASVAQAAGYGT 120  
QY 121 QPAYPAVGGQPAATATPRTPDGKPTETSQPOSSTGGYNPSLGYGSNYSYQVPGSY 180  
DB 121 QPAYPAVGGQPAATATPRTPDGKPTETSQPOSSTGGYNPSLGYGSNYSYQVPGSY 180  
QY 181 MQPVTAPSPYPTSYSTQPTSYDQSSYQONTYGGQSSYGGQSSYGGQSSYGGQSSY 240  
DB 181 MQPVTAPSPYPTSYSTQPTSYDQSSYQONTYGGQSSYGGQSSYGGQSSYGGQSSY 240  
QY 241 PPTGTSYSAQPSQYSSQSSSYGQSSYQONTYGGQSSYGGQSSYGGQSSYGGQSSY 300  
DB 241 PPTGTSYSAQPSQYSSQSSSYGQSSYQONTYGGQSSYGGQSSYGGQSSYGGQSSY 300  
QY 301 GGRGCGFDRGMRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGR 360  
DB 301 GGRGCGFDRGMRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGR 360  
QY 361 SAIVYQGLNDSVTLDDLADFFKCCGVYKAKRTGQPMIHTIYLDKXGKPKGATVSYEDP 420  
DB 361 SAIVYQGLNDSVTLDDLADFFKCCGVYKAKRTGQPMIHTIYLDKXGKPKGATVSYEDP 420  
QY 421 PTKAAVEMVFDGDFQSSKXKXVSLARKKPPMNSMRGSLPRBERGMPPRLRGPGGCGP 480  
DB 421 PTKAAVEMVFDGDFQSSKXKXVSLARKKPPMNSMRGSLPRBERGMPPRLRGPGGCGP 480  
QY 481 GGPWGMWGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGR 540  
DB 481 GGPWGMWGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGR 540  
QY 541 NCKAKYPSGFLPPPPPPGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGR 600  
DB 541 NCKAKYPSGFLPPPPPPGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGR 600  
QY 601 GGRGMRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGR 656  
DB 601 GGRGMRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGR 656

## RESULT 2

Q96FE8 PRELIMINARY; PRT; 655 AA.  
ID Q96FE8  
AC Q96FE8; 01-DEC-2001 (TEMBLrel. 19, Created)  
DT 01-DEC-2001 (TEMBLrel. 19, Last sequence update)  
DT 25-OCT-2004 (TEMBLrel. 28, Last annotation update)  
DE Ewing sarcoma breakpoint region 1, isoform EMS (EMSR1 protein).  
GN Name=EMSR1;  
OS Homo sapiens (human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
OX NCBI\_TaxId=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Lymph, and Skin;  
RX MEDLINE=23388257; PubMed=12477932; DOI=10.1073/pnas.242603699;  
RA Strausberg R.L., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,

RA Diachenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,  
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,  
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahy J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,  
RA Krzywinski M.I., Skaleja U., Smailus D.E., Scherch A., Schein J.E.,  
RA Jones S.J., Marra M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length human  
RT and mouse cDNA sequences.";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Lymph;  
RA Strausberg R.;  
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Skin;  
RA Strausberg R.;  
RL Submitted (JUN-2004) to the EMBL/GenBank/DBJ databases.  
RN [4]  
RP SEQUENCE FROM N.A.  
RA Collins J.E., Wright C.L., Edwards C.A., Davis M.P., Grinham J.A.,  
RA Cole C.G., Goward M.B., Aguado B., Mallia M., Mokrab Y., Huckle E.J.,  
RA Beare D.M., Dunham I.;  
RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.  
DR EMBL; BC011048; AAH11048.1; -;  
DR EMBL; BC072442; AAH72442.1; -;  
DR EMBL; CR456490; CAG30376.1; -;  
DR HSSP; 095218; 1N02;  
DR GO; GO:0005634; C:nucleus; IEA;  
DR GO; GO:0003676; F:nucleic acid binding; IEA;  
DR GO; GO:0008270; F:zinc ion binding; IEA;  
DR InterPro; IPR011368; RNA-binding\_EWS;  
DR InterPro; IPR000504; RNA\_rec\_mot.  
DR InterPro; IPR01876; Znf\_RangDP.  
DR Pfam; PF00641; ZF-RanBP; 1.  
DR Pfam; PF00641; ZF-RanBP; 1.  
DR SMART; SMART002101; RNA-binding\_EWS; 1.  
DR SMART; SMART00360; RRM; 1.  
DR SMART; SMART00547; ZNF\_RBZ; 1.  
DR PROSITE; PS50102; RRM; 1.  
DR PROSITE; PS01358; ZF\_RANBP2\_1; 1.  
DR PROSITE; PS50199; ZF\_RANBP2\_2; 1.  
SQ SEQUENCE 655 AA; 68391 MW; B539ED1E98C601ED CRC64;

Query Match 99.6%; Score 3618.5; DB 2; Length 655;  
Best Local Similarity 99.8%; Pred. No. 1, 1e-147;  
Matches 655; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1 MASTDSTYSQAQAAGGYSATYTAQPTGTAQTTQAVGQSSYGYGPTDVSYTAQTTAT 60  
DB 1 MASTDSTYSQAQAAGGYSATYTAQPTGTAQTTQAVGQSSYGYGPTDVSYTAQTTAT 60  
QY 61 YGQTAATATSGQPTGTTPTAQAASQPVQSGTGAVDTTTATVTTTQASVAQAAGYGT 120  
DB 61 YGQTAATATSGQPTGTTPTAQAASQPVQSGTGAVDTTTATVTTTQASVAQAAGYGT 120  
QY 121 QPAYPAVGGQPAATATPRTPDGKPTETSQPOSSTGGYNPSLGYGSNYSYQVPGSY 180  
DB 121 QPAYPAVGGQPAATATPRTPDGKPTETSQPOSSTGGYNPSLGYGSNYSYQVPGSY 180  
QY 181 MQPVTAPSPYPTSYSTQPTSYDQSSYQONTYGGQSSYGGQSSYGGQSSYGGQSSY 240  
DB 181 MQPVTAPSPYPTSYSTQPTSYDQSSYQONTYGGQSSYGGQSSYGGQSSYGGQSSY 240  
QY 241 PPTGTSYSAQPSQYSSQSSSYGQSSYQONTYGGQSSYGGQSSYGGQSSYGGQSSY 300



ID	GENVA3	PRELIMINARY:	PRT:	656 AA.
AC	GENVA3;			
DT	05-JUL-2004 (TREMBLrel. 27, Created)			
DT	05-JUL-2004 (TREMBLrel. 27, Last sequence update)			
DT	05-JUL-2004 (TREMBLrel. 27, Last annotation update)			
DE	Ewing sarcoma breakpoint region 1.			
GN	Name=Ewari;			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sclerognathi; Muridae; Murinae; Mus.			
OX	NCBI_Taxid=10090;			
RN	[1]			
RP	SEQUENCE FROM N. A.			
RC	STRAIN=C57BL/6; TISSUE=Brain;			
RX	MEDLINE=22386257; PubMed=12477932; DOI=10.1073/pnae.242603899;			
RA	Strauberg R. L., Feingold B. A., Grouse L. H., Derge J. G.,			
RA	Klausner R. D., Collins F. S., Wagner L., Shemen C. M., Schler G. D.,			
RA	Abschul S. F., Zeeberg B., Buetow K. H., Schaefer C. F., Bat N. K.,			
RA	Hopkins R. F., Jordan H., Moore T., Max S. I., Wang J., Hsieh F.,			
RA	Dichenko L., Marutina K., Farmer A. A., Rubin G. M., Hong L.,			
RA	Stapleton M., Soares M. B., Bonaldi M. F., Casavant T. L., Scheetz T. E.,			
RA	Brownstein M. J., Usdin T. B., Toshiyuki S., Carninci P., Prange C.,			
RA	Raha S. S., Loquellano N. A., Peters G. J., Abramson R. D., Mullaly S. J.,			
RA	Bosak S. A., McEwen P. J., McKernan K. J., Malak J. A., Gunaratne P. H.,			
RA	Richards S., Worley K. C., Hale S. S., Garcia A. M., Gay L. J., Hultky S. W.,			
RA	Villalón D. K., Muzny D. M., Sodergren E. J., Lu X., Gibbs R. A.,			
RA	Faney J., Helton E., Kettelman M., Madan A. C., Roddick S., Sanchez A.,			
RA	Whiting M., Madan A., Young A. C., Shevchenko Y., Bouffard G. G.,			
RA	Blakeley R. W., Touchman J. W., Green E. D., Dickson M. C.,			
RA	Rodriguez A. C., Grimwood J., Schmutz J., Myers R. M., Butlerfield Y. S.,			
RA	Krzywinski M. I., Skalske U., Smalios D. E., Schnerch A., Schein J. E.,			
RA	Jones S. U., Maira M. A.;			
RT	"Generation and initial analysis of more than 15,000 full-length human			
RT	and mouse cDNA sequences."			
RL	Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).			
RN	[2]			
RP	SEQUENCE FROM N. A.			
RC	STRAIN=C57BL/6; TISSUE=Brain;			
RC	Director MGC Project;			
RL	Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.			
DR	EMBL, BC068226, AAH68226.1; -			
DR	GO, GO:0005634; C:nucleus; IDA.			
DR	InterPro, IPR011368; RNA-binding, EMS.			
DR	InterPro, IPR000504; RNA_rec mot.			
DR	InterPro, IPR001876; znf_RanBP.			
DR	Pfam, PF00076; RRM 1; 1.			
DR	Pfam, PF00641; zf_RanBP; 1.			
DR	PIRSF, PIRSF002101; RNA-binding_EWS; 1.			
DR	SMART, SM00360; RRM; 1.			
DR	SMART, SM00547; znf_RBZ; 1.			
DR	PROSITE, PS0102; RRM; 1.			
DR	PROSITE, PS01358; zf_RanBP2_1; 1.			
DR	PROSITE, PS0199; zf_RanBP2_2; 1.			
SO	SEQUENCE 656 AA; 68549 MW; 5DF018A22C08C1D0B CRC64;			
Query Match	98.5%; Score 3579; DB 2; Length 656;			
Best Local Similarity	97.9%; Pred. No. 5; 6e-146;			
Matches 642; Conservative 9; Mismatches 5; Indels 0; Gaps 0				
QY	1 MASTDYSTGYSAAMAAQGGYSAYTAQPTGQYAGTTQAYVGGQSYGTGYGQPTDVSYTQAQTAT 60			
DB	1 MASTDYSTGYSAAMAAQGGYSAYTAQPTGQYAGTTQAYVGGQSYGTGYGQPTDVSYTQAQTAT 60			
QY	61 YGQRAVAATSYGQPPGYGTPPAPOAYSGQPVGCGAGADTTTATTTTQASAAAGSAYGT 120			
DB	61 YGQRAVAATSYGQPPGYGTPPAPOAYSGQPVGCGAGADTTTATTTTQASAAAGSAYGT 120			
QY	121 QPAPVAYGQGPAAATPATRPODGNKRTETSSOPSSSTGCGNNOPSLAGGGSNYSYPVPGSAYP 180			
DB	121 QPAPVAYGQGPAAATPATRPODGNKRAETSSOPSSSTGCGNNPBLGCGSNYSYPVPGSAYP 180			
QY	181 MQPVTAPEPTSYSTSTQPTSTYDQSSYSQONTYGPSSYGGQSSYGGQSSYGGQPPSY. 240			

Db	181	MQPTTAPPSYPTSYSSSSQPTSYDYSSYSQNTTTCGSSSYGQSSYQGSSTYGGQPTSTY	240
Qy	241	PPQTGSYSQABSYQSSSSSYGQSSFRQDHPSSSMGYVQGESGGFSGPGENRSMSPDNR	300
Db	241	PPQTGSYSQABSYQSSSSSYGQSSFRQDHPSSSMGYVQGESGGFSGPGENRSMSPDNR	300
Qy	301	GRGGRGPDRCGMSNGSGRGGCGMGSGAGEGRPFKPGCPMDDEGDDLGLPVPVDEDSDN	360
Db	301	GRGGRGPDRCGMSNGSGRGGCGMGSGAGEGRPFKPGCPMDDEGDDLGLPVPVDEDSDN	360
Qy	361	SAIYVQGLNDSVTLLDLADFPKQCGVYVMNRKTCQPMIHYLDKETGPKKDATVSYEDP	420
Db	361	SAIYVQGLNDSVTLLDLADFPKQCGVYVMNRKTCQPMIHYLDKETGPKKDATVSYEDP	420
Qy	421	PTAAVAEWFDPGDKFOGSKLVSLARKKPPMNSMRGGLPFRBGRGMPPLRGSGPGGP	480
Db	421	PTAAVAEWFDPGDKFOGSKLVSLARKKPPMNSMRGGLPFRBGRGMPPLRGSGPGGP	480
Qy	481	GGPMGRNGRGGDGGFPFRPGRSSRGNPSSGGGVORPADWQCPNPGCCGNQNAARTEC	540
Db	481	GGPMGRNGRGGDGGFPFRPGRSSRGNPSSGGGVORPADWQCPNPGCCGNQNAARTEC	540
Qy	541	NQCAAPPEBGLPPPPPPPGGDRGRGGPCGMRGGRGGLMDRGCGGMFRGGSGGDRGGR	600
Db	541	NQCAAPPEBGLPPPPPPPGGDRGRGGPCGMRGGRGGLMDRGCGGMFRGGSGGDRGGR	600
Qy	601	GGRGMDRGCFGGGRRGPGCGPCCPLMEQMGRRGCGRGPCKMDGGEHROERDRPY	656
Db	601	GGRGMDRGCFGGGRRGPGCGPCCPLMEQMGRRGCGRGPCKMDGGEHROERDRPY	656

RESULT 5

EWS\_MOUSE

ID	EWS_MOUSE	STANDARD;	PRT;	655 AA.
AC	061545;			
DT	01-NOV-1997 (Rel. 35, Created)			
DT	01-NOV-1997 (Rel. 35, Last sequence update)			
DT	25-OCT-2004 (Rel. 45, Last annotation update)			
DE	RNA-binding protein EWS.			
GN	Name=Ewsi; Synonym=Ews; Ewsh;			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
OX	NCBI_TaxID=10090;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Testis;			
FX	MEDLINE=95130099; PubMed=7829090;			
RA	Plougastel B., Mattei M.-G., Thomas G., Delattre O.;			
RT	"Cloning and chromosome localization of the mouse Ews gene.";			
RL	Genomics 23:278-281(1994).			
CC	-1- FUNCTION: Might function as a repressor (By similarity).			
CC	-1- SUBUNIT: Binds RNA. POLR2C, SFI and calmodulin. Interacts with			
CC	PRXB2 (By similarity).			
CC	-1- SUBCELLULAR LOCATION: Nuclear, cytoplasmic and on cell surface.			
CC	Relocates from cytoplasm to ribosomes upon PRXB2/FAK2 activation			
CC	(By similarity).			
CC	-1- PM: Phosphorylated; calmodulin-binding inhibits phosphorylation			
CC	of Ser-266 (By similarity).			
CC	-1- MISCELLANEOUS: Binds calmodulin in the presence, but not in the			
CC	absence, of calcium ion (By similarity).			
CC	-1- SIMILARITY: Belongs to the RNP TET family.			
CC	-1- SIMILARITY: Contains 1 IQ domain.			
CC	-1- SIMILARITY: Contains 1 RanzB-type zinc finger.			
CC	-1- SIMILARITY: Contains 1 RNA recognition motif (RRM) domain.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>			

or send an email to license@sb-sib.ch).

CC -----

DR EMBL: X79233; CA55815.1; -

DR PIR: A55726; A55726.

DR HSSP: 095218; INO2.

DR MGD; MGI:99960; Ewsh.

DR GO: GO:0005634; C:nucleus; IDA.

DR InterPro: IPR000048; IQ\_region.

DR InterPro: IPR000504; RNA\_rec\_mot.

DR InterPro: IPR001876; Znf\_RandP.

DR Pfam: PF00076; RRM\_1; 1.

DR Pfam: PF00641; zf\_RanBP; 1.

DR PIRSF: PIRSF002101; RNA-binding\_EWS; 1.

DR PROSITE: PS50096; IQ; FALSE\_NEG.

DR PROSITE: PS0102; RRM; 1.

DR PROSITE: PS01358; ZF\_RANBP2\_1; 1.

DR PROSITE: PS50199; ZF\_RANBP2\_2; 1.

DR Calmodulin-binding; Metal-binding; Methylation; Nuclear protein;

KM Phosphorylation; Repeat; Repressor; RNA-binding;

KM Transcription regulation; zinc; zinc-finger.

FT DOMAIN 1 285 EAD (Gln/Pro/Thr-rich).

FT DOMAIN 256 285 IQ.

FT DOMAIN 360 446 RNA-binding (RRM).

FT ZN\_FING 517 548 RanBP2-type.

FT 8 285 31 X approximate tandem repeats.

FT REPEAT 17 27 2.

FT REPEAT 28 34 3.

FT REPEAT 35 42 4.

FT REPEAT 43 50 4.

FT REPEAT 51 59 6.

FT REPEAT 60 68 7.

FT REPEAT 69 75 8.

FT REPEAT 76 84 9.

FT REPEAT 85 91 10.

FT REPEAT 92 110 11.

FT REPEAT 111 116 12.

FT REPEAT 117 125 13.

FT REPEAT 126 156 14.

FT REPEAT 157 163 15.

FT REPEAT 164 170 16.

FT REPEAT 171 177 17.

FT REPEAT 178 188 18.

FT REPEAT 189 193 19.

FT REPEAT 194 201 20.

FT REPEAT 202 206 21.

FT REPEAT 207 212 22.

FT REPEAT 213 218 23.

FT REPEAT 219 224 24.

FT REPEAT 225 230 25.

FT REPEAT 231 238 26.

FT REPEAT 239 245 27.

FT REPEAT 246 252 28.

FT REPEAT 253 259 29.

FT REPEAT 260 276 30.

FT REPEAT 277 285 31.

FT DOMAIN 300 339 Arg/Gly/Pro-rich.

FT DOMAIN 453 512 Arg/Gly/Pro-rich.

FT DOMAIN 558 639 Arg/Gly/Pro-rich.

FT MOD\_RES 300 Asymmetric dimethylarginine (By similarity).

FT MOD\_RES 302 Asymmetric dimethylarginine (By similarity).

FT MOD\_RES 304 Asymmetric dimethylarginine (By similarity).

FT MOD\_RES 309 Asymmetric dimethylarginine (By similarity).

FT MOD\_RES 314 Asymmetric dimethylarginine (By similarity).

FT MOD\_RES 317 Asymmetric dimethylarginine (By similarity).

FT MOD\_RES 321 Asymmetric dimethylarginine (By similarity).

FT MOD\_RES 454 454 Asymmetric dimethylarginine (By similarity).

FT MOD\_RES 463 463 Asymmetric dimethylarginine (By similarity).

FT MOD\_RES 470 470 Asymmetric dimethylarginine (By similarity).

FT MOD\_RES 489 489 Asymmetric dimethylarginine (By similarity).

FT MOD\_RES 493 493 Asymmetric dimethylarginine (By similarity).

FT MOD\_RES 499 499 Asymmetric dimethylarginine (By similarity).

FT MOD\_RES 502 502 Asymmetric dimethylarginine (By similarity).

FT MOD\_RES 505 505 Asymmetric dimethylarginine (By similarity).

FT MOD\_RES 562 562 Asymmetric dimethylarginine (By similarity).

FT MOD\_RES 564 564 Asymmetric dimethylarginine (By similarity).

FT MOD\_RES 571 571 Asymmetric dimethylarginine (By similarity).

FT MOD\_RES 574 574 Asymmetric dimethylarginine (By similarity).

FT MOD\_RES 580 580 Asymmetric dimethylarginine (By similarity).

FT MOD\_RES 588 588 Asymmetric dimethylarginine (By similarity).

FT MOD\_RES 591 591 Asymmetric dimethylarginine (By similarity).

FT MOD\_RES 595 595 Asymmetric dimethylarginine (By similarity).

FT MOD\_RES 599 599 Asymmetric dimethylarginine (By similarity).

FT MOD\_RES 602 602 Asymmetric dimethylarginine (By similarity).

FT MOD\_RES 606 606 Asymmetric dimethylarginine (By similarity).

FT MOD\_RES 614 614 Asymmetric dimethylarginine (By similarity).

FT MOD\_RES 632 632 Asymmetric dimethylarginine (By similarity).

FT MOD\_RES 635 635 Asymmetric dimethylarginine (By similarity).

FT MOD\_RES 655 AA; 68418 MW; 50735EDB54247D69 CRC64; Phosphoserine (By PKC) (By similarity).

SO SEQUENCE

Query Match 98.3%; Score 3571.5; DB 1; Length 655;

Best Local Similarity 98.0%; Pred. No. 1.2e-145;

Matches 643; Conservative 8; Mismatches 4; Indels 1; Gaps 1;

QY 1 MASTDSTYSQAQAQGSAYTAOPTQGAOTTAQAGQSSYGVGQPTDVSYQAQTTAT 60

DB 1 MASTDSTYSQAQAQGSAYTAOPTQGAOTTAQAGQSSYGVGQPTDVSYQAQTTAT 60

QY 61 YGQTAAYATSYGQPTGYTTPTAQAAYSQPVQGYGTGAVDTTATVTTTQAASVAAGAT 120

DB 61 YGQTAAYATSYGQPTGYTTPTAQAAYSQPVQGYGTGAVDTTATVTTTQAASVAAGAT 120

QY 121 QPAYPAYGQOPATATPATRQDQNKPTETQPOSSTGYNQPSLGYQGSYNTSYQVPGSTP 180

DB 121 QPAYPAYGQOPATATPATRQDQNKPTETQPOSSTGYNQPSLGYQGSYNTSYQVPGSTP 180

QY 181 MOPVTAPEPYPTPSYSTQPTSDSSYQONTYQPPSSYGQSSYGGQSSYVQGPPTSY 240

DB 181 MOPVTAPEPYPTPSYSTQPTSDSSYQONTYQPPSSYGQSSYGGQSSYVQGPPTSY 240

QY 241 PPGTGSYQAPQYSGQSSYQSSYFRQDHSMSGVYQGESGFGSPENRSMGSPDNR 300

DB 241 PPGTGSYQAPQYSGQSSYQSSYFRQDHSMSGVYQGESGFGSPENRSMGSPDNR 300

QY 301 GKGGGFDRGMSRGRGGGGRGMSAGRGGFGKPGGPMDEGPDLDLGPVDPDESDN 360



Db	301	GRGRGGRFPDRGCMGSRGRGGRGGRGGLG-AGERGCGRNKGCPMPDEBPDLGLPIPDDESDN	353
Qy	361	SAIYYQGLNDSVTLDDLADFPKQCGVYKXNKRTGTGPIMHIYLDKETGKPKGATVSYEDP	420
Db	360	SAIYYQGLINDNTVTLDDLADFPKQCGVYKXNKRTGTGPIMHIYLDKETGKPKGATVSYEDP	41.9
Qy	421	PTAKAAVEMFQCKDDQGSGLXYSLAKKPKPMNSMRGGLPPREBERGMPPLLRGPGPGGPG	480
Db	420	PTAKAAVEFPDKDDQGSGLXYSLAKKPKPMNSMRGGLPPREBERGMPPLLRGPGPGGPGGPG	479
Qy	481	GGPMRMGCGRGDRGCFPRGPRGRSGNSGGGCVYHRAAGDMQCPNPGCGNPFARTEC	540
Db	480	GGPMRMGCGRGDRGCFPRGPRGRSGNSGGGCVYHRAAGDMQCPNPGCGNPFARTEC	539
Qy	541	NOCKAPKEGFLPPPPPPGDRGGRGPGCGMRGGRGLMDRGPGGMPFRGGRGDRGCFR	600
Db	540	NOCKAPKEGFLPPPPPPGDRGGRGPGCGMRGGRGLMDRGPGGMPFRGGRGDRGCFR	599
Qy	601	GGRGMDRGFGGCGRRGCGPGGPPGGLMEQMGRRGGGCGKXDKGHRDRRDRPY	656
Db	600	GGRGMDRGFGGCGRRGCGPGGPPGGLMEQMGRRGGGCGKXDKGHRDRRDRPY	655

RESULT 6

ID	Q9D2P0	PRELIMINARY;	PRT;	655 AA.
AC	Q9D2P0;			
DT	01-JUN-2001 (TREMBLrel. 17, Created)			
DT	01-JUN-2001 (TREMBLrel. 17, Last sequence update)			
DT	01-MAR-2004 (TREMBLrel. 26, Last annotation update)			
DE	Mus musculus 18 days pregnant adult female placenta and extra embryonic tissue cDNA, RIKEN full-length enriched library, clone:1830417B11 product:Ewing sarcoma homolog, full insert sequence.			
DE	GN Name:Ewari; Synonym:Ewiah;			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Cranata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
OX	NCBI_Taxid:10090;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=C57BL/6J; TISSUE=Placenta and extra embryonic tissue;			
RC	MEDLINE=21085660; PubMed=10345636; DOI=10.1016/S0076-6879(99)03004-9;			
RA	RIKEN FANTOM Consortium;			
RT	"Functional annotation of a full-length mouse cDNA collection.";			
RL	Nature 403:685-690(2001).			
RN	[3]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=C57BL/6J; TISSUE=Placenta and extra embryonic tissue;			
RC	MEDLINE=20499374; PubMed=11042159; DOI=10.1101/9T.145100;			
RA	Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M., Kono H., Okazaki Y., Muramatsu M., Hayashizaki Y.;			
RT	"Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes.";			
RL	Genome Res. 10:1617-1630(2000).			
RN	[5]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=C57BL/6J; TISSUE=Placenta and extra embryonic tissue;			

Query Match	Best Local Similarity	98.1%	Score 3564.5	DB 2	Length 655				
Matches	641	Conservative	9	Mismatches	5	Indels	1	Gaps	1
QY	1	MASIDYSTYSGAAAGGYSAYTAQPTGQYAGTQAAGQSSYGTGQPTDVSYTAQTTAT	60						
DB	1	MASIDYSTYSGAAAGGYSAYTAQPTGQYAGTQAAGQSSYGTGQPTDVSYTAQTTAT	60						
QY	61	YGGTAATSYGQPTGTYTTPAIPAQYISQPVGTCGAYDTTATYTTTQASAAQSAVGT	120						
DB	61	YGGTAATSYGQPTGTYTTPAIPAQYISQPVGTCGAYDTTATYTTTQASAAQSAVGT	120						
QY	121	QAPATPAAGQQAATPAATRPDQGNKPTESQPSQSSGTCGNQPSLIGQSSNYSPOVPGSY	180						
DB	121	QAPATPAAGQQAATPAATRPDQGNKPTESQPSQSSGTCGNQPSLIGQSSNYSPOVPGSY	180						
QY	181	MQPYTAPPSYPTSYSSSTOPTSYDQSSYSGQNTTGPSSYGGQSSYGGQPPSY	240						
DB	181	MQPYTAPPSYPTSYSSSTOPTSYDQSSYSGQNTTGPSSYGGQSSYGGQPPSY	240						
QY	241	PPQYGSYSQAPSYQSSQSSSYGQSSSFRODHPSSMGVYGQSSGSGPGENRSLSGPNNR	300						
DB	241	PPQYGSYSQAPSYQSSQSSSYGQSSSFRODHPSSMGVYGQSSGSGPGENRSLSGPNNR	300						
QY	301	GRGGGGRDRGMSGGGGGGGSGAGERGPFNKGPMGEGPDLIGPVPDPEDSDN	360						
DB	301	GRGGGGRDRGMSGGGGGGGSGAGERGPFNKGPMGEGPDLIGPVPDPEDSDN	360						
QY	361	SAIVVGLNDSVTLDDLADFPKQCGVVMNRGTQPMHLYLDKETGPKKDAATVSYDP	420						
DB	361	SAIVVGLNDSVTLDDLADFPKQCGVVMNRGTQPMHLYLDKETGPKKDAATVSYDP	420						

Db	360	SALVYQGLNDVNTLDDLDADFFKQCCGVVKNKKRTQCEMIH1YLDKETGKPEKDATVSYBDP	41
Qy	421	PTAKAAVEMPPDCKDQVGSGLTKYSLARKKRPNNMSMRGLPPRBERGNMPPPLRGPGPGGP	480
Db	420	PTAKAAVEFPDCKDQVGSGLTKYSLARKKRPNNMSMRGMPPRBERGNMPPPLRGPGPGGP	479
Qy	481	GGPMRMGRCGRDRCGFPFPRGRCGRSGKNSGGGVNHRAGDMQCPNPGCGNQPARTTEC	540
Db	480	GGPMRMGRCGRDRCGFPFPRGRCGRSGKNSGGGVNHRAGDMQCPNPGCGNQPARTTEC	539
Qy	541	NOCKAPKCEGFLPPPPPPPGDRCGRGCGPGGMRCGLMDRGPGGMFRGRCGRDRCGR	600
Db	540	NOCKAPKCEGFLPPPPPPPGDRCGRGCGPGGMRCGRGGLMDRGPGGMFRGRCGRDRCGR	599
Qy	601	GGRGMDRCGFGGRCGRGCGPGGPQPLMEQNGRCRRGCGGCGIXMDKGEHROBRARPY	656
Db	600	GGRGMDRCGFGGRCGRGCGPGGPQPLMEQNGRCRRGCGGCGIXMDKGEHROBRARPY	655
RESULT 7			
Q96MN4	096MN4	PRELIMINARY; PRT; 600 AA.	
AC	096MN4		
DT	01-DEC-2001 (TrEMBLrel. 19, Created)		
DT	01-DEC-2001 (TrEMBLrel. 19, Last sequence update)		
DT	01-MAR-2004 (TrEMBLrel. 26, Last annotation update)		
DE	Hypothetical protein FLJ32119.		
OS	Homo sapiens (human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
OX	NCBI_TaxId=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	PubMed=14702039; DOI=10.1038/ng1285;		
RA	Oca T., Suzuki Y., Nishikawa T., Otsuki T., Sugiyama T., Irie R.,		
RA	Wakamatsu A., Hayashi K., Sato H., Nagai K., Kimura K., Makita H.,		
RA	Sekine M., Ohyashi M., Nishi T., Shiohara T., Tanaka T., Ishii S.,		
RA	Yanamoto J., Saito K., Kawai Y., Isono Y., Nakamura Y., Negahari K.,		
RA	Murakami K., Yasuda T., Iwayanagi T., Wagatsuma M., Shiratori A.,		
RA	Sudo H., Hosoiri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,		
RA	Takahashi M., Kanda K., Yokoi T., Furuya T., Kikkawa E., Omura Y.,		
RA	Abe K., Kamihara K., Katsura N., Suto K., Tanikawa M., Yamazaki M.,		
RA	Ninomiya K., Ishibashi T., Yamashita K., Murakawa K., Fujimori S.,		
RA	Tanai H., Kimeta M., Watanabe M., Hirooka S., Chiba Y., Ishida S.,		
RA	Ono Y., Takiguchi S., Watanabe S., Yoshida M., Houcha T., Kusano J.,		
RA	Kanehori K., Takahashi-Fujii A., Hara H., Tanase T., Nomura Y.,		
RA	Togiyama S., Komai F., Hara R., Takeuchi K., Arita M., Imose N.,		
RA	Yoshikawa K., Yunki H., Oshima A., Saeki N., Aotsuka S.,		
RA	Moriya S., Moriyama H., Ichihara T., Shiohara N., Sano S.,		
RA	Nishigawa S., Senoh A., Mizoguchi H., Goto Y., Shimizu F., Wakebe H.,		
RA	Hibishigaki H., Watanabe T., Sugiyama A., Takemoto M., Kawakami B.,		
RA	Fujimaki M., Watanabe K., Kumagai A., Itakura S., Fukuzumi Y.,		
RA	Fujimori Y., Komiyama M., Tashiro H., Tanigami A., Fujitawa T.,		
RA	Ono T., Yamada K., Fujii Y., Ozaki K., Hirao M., Ohmori Y.,		
RA	Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,		
RA	Okitani R., Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T.,		
RA	Matsumura K., Nakajima Y., Mizuno T., Morinaga M., Saeki M.,		
RA	Togehashi T., Oyama M., Hata H., Watanabe M., Komatsu T.,		
RA	Mizushima-Sugano J., Sato T., Shirai Y., Takahashi Y., Nakagawa K.,		
RA	Okumura K., Nagase T., Nomura N., Kikuchi H., Masuho Y., Yamashita R.,		
RA	Nakai K., Yada T., Nakamura Y., Ohara O., Isegaki T., Sugano S.,		
RT	"Complete sequencing and characterization of 21,243 full-length human		
RT	cDNAs";		
RL	Nat. Genet. 36:40-45(2004).		
DR	EMBL; AK056681; BAB71252.1; -.		
DR	HSSP; 095218; INOZ.		
DR	GO; GO:0005634; C:nucleus; IEA.		
DR	GO; GO:0003676; P:nucleic acid binding; IEA.		
DR	GO; GO:0008270; P:zinc ion binding; IEA.		
DR	InterPro; IPR011368; RNA-binding_EMS.		
DR	InterPro; IPR000504; RNA_rec_moc.		
DR	InterPro; IPR001876; ZnF_RanGDP.		

[illegible]

RESULT 9

```
Q6DJ04 PRELIMINARY;  PRT;  673 AA.
Q6DJ04
AC Q6DJ04;
DT 25-OCT-2004 (Tremblrel. 28, Created)
DT 25-OCT-2004 (Tremblrel. 28, Last sequence update)
DT 25-OCT-2004 (Tremblrel. 28, Last annotation update)
DE Hypothetical protein MGC76258.
GN Name=MGC76258;
OS Xenopus tropicalis (Western clawed frog) (Silurana tropicalis).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8364;
[1]
RN
RP SEQUENCE FROM N.A.
RC TISSUE=Embryo;
RX MEMLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Haile F.,
RA Diatchenko L., Marsina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stepieton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein W.J., Udell T.B., Toshilyki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
RA Bosak S.A., McGowan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Huliy S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton B., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywnicki M.I., Skalka U., Smailus D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences."
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
[2]
RN
RP SEQUENCE FROM N.A.
RC TISSUE=Embryo;
RA Klein S., Gerhard D.S.;
RL Submitted (JUN-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC075120; AAH75120.1;
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR GO; GO:0008270; F:zinc ion binding; IEA.
DR Interpro; IPR006162; Ppantne S.
DR Interpro; IPR011368; RNA-binding EMS.
```



```

QY 228 QGSSY-----GGQPTSYSPONGSYSQ-APSOYSGQSSS-----YGOQ-----SSFR 268
Db 197 QGCGYQOTPPQQQABPPSSYAPSSGSGYGSAGYGGQSTGGGYSYKPPNQGSYR 256
QY 269 QDHPSSMGVYQBSGGFSGPGENRSMSPNRRGGRGPGFPGGMSRGG--RGG-GRGGMG 325
Db 257 PDHONGGGVYGGPSSGGYGGPGEGRMGCGENRGGRGGRGPGFPGGMRGGCGGRGMSRGGMG 316
QY 326 SAGEGPGFNRKPGGPMDEGPDLDLGPDPDPEDSDNSAIYVQGLNDSYTLDDLADFPKQ 385
Db 317 IADRGGPGFNRKPG-----DGEAGPAREOD--DSENSITYTGLTENATLEEVDFPFHSG 368
QY 386 VVGNKRTGQPMHIYLDKETGPKGPDATSYEDPPAKAAVMPDGKDFGSKLXSLA 445
Db 369 IIRINRTGIPAVNIYTDKDTGPKGPDATSYEPPEAKAAVEMFDGDKDQGGKTLXSLA 428
QY 446 RKPPNMSNRGGLPPRGRGMPPLRGRGPGGPGGPGMGRMG-GRGDRGCGFPPR-GPR 503
Db 429 RRKPMNGMNRGGM-----PMRGDRGMMGRGGMGRGGMGRGGRGDRGGMPPRGGR 478
QY 504 G-SRGNPSSGGANVOHRAQDMQCPNPGCGNPFAMRTCCNOCKAPKPEGFLPPPPPPG 562
Db 479 GMRGGRGT--GNNQOQRAQDMQCPNAGCGNPFAMRMECNQCKAPKPEGFPFPPF-PFG 536
QY 563 RGRGPGGMRGGRGLMDRGGPGGM--FRGGRGDRGGRGGRGMRGGRGGRGGRGPG 620
Db 537 RGRGPGGMRGGRG--MDRGGPGGPGGFRGGRGDRGGRGFP--GRGMDGFGG--RGRGG 590
QY 621 PPGPLMOMGRRGRGGRGPGKMD-KGEHROERRDRPY 656
Db 591 PF---MDMGRRGRGMRGMPGCKMDKGDHRRDRERY 624

RESULT 11
Q9CRSS PRELIMINARY; PRT; 333 AA.
AC Q9CR85;
DT 01-JUN-2001 (TEMBLrel. 17, Created)
DT 01-JUN-2001 (TEMBLrel. 17, Last sequence update)
DT 01-MAR-2004 (TEMBLrel. 26, Last annotation update)
DE Mus musculus 17 days embryo head cDNA, RIKEN full-length enriched
DE library, clone:330002D11 product:Ewing sarcoma homolog, full insert
DE sequence. (fragment).
GN Name=Bwer1, Syntonym=Bwh1;
OS Mus musculus (Mouse);
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Head;
RC MEDLINE=9279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;
RA Carninci P., Hayashizaki Y.;
RT "High-efficiency full-length cDNA cloning.";
RL Mech. Enzymol. 303:19-44(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Head;
RC MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;
RA RIKEN PANTOM Consortium;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Head;
RA The PANTOM Consortium.
RA The RIKEN Genome Exploration Research Group Phase I & II Team.
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573(2002).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Head;

```

```

RX MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;
RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,
RA Kono H., Okazaki Y., Muramatsu M., Hayashizaki Y.;
RT "Normalization and subtraction of cap-trapper-selected cDNAs to
RT prepare full-length cDNA libraries for rapid discovery of new genes.";
RL Genome Res. 10:1617-1630(2000).
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Head;
RC MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;
RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,
RA Kono H., Akiyama J., Nishi K., Katsunai T., Tashiro H., Itoh M.,
RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,
RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
RA Fujiwara S., Inoue K., Togawa K., Tanaka T., Matsumura S., Kawai J.,
RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsumura S., Kawai J.,
RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
RT "RIKEN integrated sequence analysis (RISA) system-384-format
RT sequencing pipeline with 384 multicapillary sequencer.";
RL Genome Res. 10:11757-1771(2000).
RN [6]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Head;
RA Adachi J., Aizawa K., Akahira S., Akimura T., Arai A., Aono H.,
RA Arakawa T., Bono H., Carninci P., Fukuda S., Fukunishi Y., Furuno M.,
RA Hanagaki T., Hara A., Hayatsu N., Hiramoto K., Hirata T., Hori F.,
RA Iwamoto K., Ishii Y., Itoh M., Izawa M., Kasakawa T., Kato H.,
RA Kawai J., Kojima Y., Kono H., Kouda M., Koya S., Kurihara C.,
RA Matsuyama T., Miyazaki A., Nishi K., Nomura K., Numazaki R., Ohno M.,
RA Okazaki Y., Okido T., Owa C., Saito H., Saito R., Sakai C., Sakai K.,
RA Sano H., Sasaki D., Shibata K., Shibata Y., Shingawa A., Shiraki T.,
RA Sogabe Y., Suzuki H., Tagami M., Tagawa A., Takahashi F., Tanaka T.,
RA Tejima Y., Toya T., Yamamura T., Yasunishi A., Yoshida K., Yoshino M.,
RA Muramatsu M., Hayashizaki Y.;
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK014366; BAB29301.1; -.
DR HSSP; O95218; 1N02.
DR MCD; MG1:99960; Bwer1.
DR GO; GO:0005634; C:nucleus; IDA.
DR InterPro; IPR000504; RNA_rec_mot.
DR InterPro; IPR001876; Znf_RanGDP.
DR Pfam; PF00076; RRM_1; 1.
DR Pfam; PF00641; Zf_RanBP; 1.
DR SMART; SM00360; RRM; 1.
DR SMART; SM00547; Znf_RBZ; 1.
DR PROSITE; PS0102; RRM; 1.
DR PROSITE; PS01358; ZF_RANBP2_1; 1.
DR PROSITE; PS50199; ZF_RANBP2_2; 1.
FT NON_TER 1 1
SQ SEQUENCE 333 AA; 34617 MW; CAEDCA353FF2P07 CRC64;

Query Match 51.6%; Score 1874; DB 2; Length 333;
Best Local Similarity 98.5%; Pred. No. 3,1e-73;
Matches 328; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 324 MSGAGRGGPNKPGGPMDEGPDLDLGPDPDPEDSDNSAIYVQGLNDSYTLDDLADFPKQ 383
Db 1 LGSAGRGGPNKPGGPMDEGPDLDLGPDPDPEDSDNSAIYVQGLNDSYTLDDLADFPKQ 60
QY 384 CGVYKNNKRTGQPMHIYLDKETGPKGPDATSYEDPPAKAAVEMFDGDKDFGSKLXYS 443
Db 61 CGVYKNNKRTGQPMHIYLDKETGPKGPDATSYEDPPAKAAVEMFDGDKDFGSKLXYS 120
QY 444 LARKKPPNMSNRGGLPPRGRGMPPLRGRGPGGPGGPGMGRMGGRGDRGGRGFPFPPRGR 503
Db 121 LARKKPPNMSNRGGMPPRGRGMPPLRGRGPGGPGGPGMGRMGGRGDRGGRGFPFPPRGR 180
QY 504 GSRGNPSSGGANVOHRAQDMQCPNPGCGNPFAMRTCCNOCKAPKPEGFLPPPPPPG 563
Db 181 GSRGNPSSGGANVOHRAQDMQCPNPGCGNPFAMRTCCNOCKAPKPEGFLPPPPPPG 240
QY 564 GRGGPGGMRGGRGLMDRGGPGGM--FRGGRGDRGGRGGRGMRGGRGGRGGRGPG 623

```

Db	Query	Db	Query	Score	Length	DB 2	Length	DB 3	Length	DB 4	Length	DB 5	Length	DB 6	Length	DB 7	Length	DB 8	Length	DB 9	Length	DB 10	Length	DB 11	Length	DB 12	Length	DB 13	Length	DB 14	Length	DB 15	Length	DB 16	Length	DB 17	Length	DB 18	Length	DB 19	Length	DB 20	Length	DB 21	Length	DB 22	Length	DB 23	Length	DB 24	Length	DB 25	Length	DB 26	Length	DB 27	Length	DB 28	Length	DB 29	Length	DB 30	Length	DB 31	Length	DB 32	Length	DB 33	Length	DB 34	Length	DB 35	Length	DB 36	Length	DB 37	Length	DB 38	Length	DB 39	Length	DB 40	Length	DB 41	Length	DB 42	Length	DB 43	Length	DB 44	Length	DB 45	Length	DB 46	Length	DB 47	Length	DB 48	Length	DB 49	Length	DB 50	Length	DB 51	Length	DB 52	Length	DB 53	Length	DB 54	Length	DB 55	Length	DB 56	Length	DB 57	Length	DB 58	Length	DB 59	Length	DB 60	Length	DB 61	Length	DB 62	Length	DB 63	Length	DB 64	Length	DB 65	Length	DB 66	Length	DB 67	Length	DB 68	Length	DB 69	Length	DB 70	Length	DB 71	Length	DB 72	Length	DB 73	Length	DB 74	Length	DB 75	Length	DB 76	Length	DB 77	Length	DB 78	Length	DB 79	Length	DB 80	Length	DB 81	Length	DB 82	Length	DB 83	Length	DB 84	Length	DB 85	Length	DB 86	Length	DB 87	Length	DB 88	Length	DB 89	Length	DB 90	Length	DB 91	Length	DB 92	Length	DB 93	Length	DB 94	Length	DB 95	Length	DB 96	Length	DB 97	Length	DB 98	Length	DB 99	Length	DB 100	Length	DB 101	Length	DB 102	Length	DB 103	Length	DB 104	Length	DB 105	Length	DB 106	Length	DB 107	Length	DB 108	Length	DB 109	Length	DB 110	Length	DB 111	Length	DB 112	Length	DB 113	Length	DB 114	Length	DB 115	Length	DB 116	Length	DB 117	Length	DB 118	Length	DB 119	Length	DB 120	Length	DB 121	Length	DB 122	Length	DB 123	Length	DB 124	Length	DB 125	Length	DB 126	Length	DB 127	Length	DB 128	Length	DB 129	Length	DB 130	Length	DB 131	Length	DB 132	Length	DB 133	Length	DB 134	Length	DB 135	Length	DB 136	Length	DB 137	Length	DB 138	Length	DB 139	Length	DB 140	Length	DB 141	Length	DB 142	Length	DB 143	Length	DB 144	Length	DB 145	Length	DB 146	Length	DB 147	Length	DB 148	Length	DB 149	Length	DB 150	Length	DB 151	Length	DB 152	Length	DB 153	Length	DB 154	Length	DB 155	Length	DB 156	Length	DB 157	Length	DB 158	Length	DB 159	Length	DB 160	Length	DB 161	Length	DB 162	Length	DB 163	Length	DB 164	Length	DB 165	Length	DB 166	Length	DB 167	Length	DB 168	Length	DB 169	Length	DB 170	Length	DB 171	Length	DB 172	Length	DB 173	Length	DB 174	Length	DB 175	Length	DB 176	Length	DB 177	Length	DB 178	Length	DB 179	Length	DB 180	Length	DB 181	Length	DB 182	Length	DB 183	Length	DB 184	Length	DB 185	Length	DB 186	Length	DB 187	Length	DB 188	Length	DB 189	Length	DB 190	Length	DB 191	Length	DB 192	Length	DB 193	Length	DB 194	Length	DB 195	Length	DB 196	Length	DB 197	Length	DB 198	Length	DB 199	Length	DB 200	Length	DB 201	Length	DB 202	Length	DB 203	Length	DB 204	Length	DB 205	Length	DB 206	Length	DB 207	Length	DB 208	Length	DB 209	Length	DB 210	Length	DB 211	Length	DB 212	Length	DB 213	Length	DB 214	Length	DB 215	Length	DB 216	Length	DB 217	Length	DB 218	Length	DB 219	Length	DB 220	Length	DB 221	Length	DB 222	Length	DB 223	Length	DB 224	Length	DB 225	Length	DB 226	Length	DB 227	Length	DB 228	Length	DB 229	Length	DB 230	Length	DB 231	Length	DB 232	Length	DB 233	Length	DB 234	Length	DB 235	Length	DB 236	Length	DB 237	Length	DB 238	Length	DB 239	Length	DB 240	Length	DB 241	Length	DB 242	Length	DB 243	Length	DB 244	Length	DB 245	Length	DB 246	Length	DB 247	Length	DB 248	Length	DB 249	Length	DB 250	Length	DB 251	Length	DB 252	Length	DB 253	Length	DB 254	Length	DB 255	Length	DB 256	Length	DB 257	Length	DB 258	Length	DB 259	Length	DB 260	Length	DB 261
----	-------	----	-------	-------	--------	------	--------	------	--------	------	--------	------	--------	------	--------	------	--------	------	--------	------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	-------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------

Dc		2	MASADYNSYGAGSGGCGXGYPAQPSQPGVGGCTQQSYGQQQGSYSVPADSTYSQT5	61
Qy		57	-TTATYGTGAATATSTGCGPPT---GTYTTLA-PLAYSQPVGCGGAAIDTTTA---VTYTT	108
Dc		62	GSSSSYGGQPYSSSYGGQPPPASGASNASAAPFGYTGPVQGYSSTSDSSTAASSTSNT	121
Qy		109	OASVYAQSAYGTQPPAYPAVGGOOPALATPRTPDQDNKPLETETSQPOSSTGCYNOPSIGYQS	168
Dc		122	QTSYGGQASAYGTQSAIPGEGQDPAAAAP-----PESTYSSSGQQA-greos	166
Qy		169	NYSYPQVPSSYPMQGVTPATPASTYPTSTYSSTOPTSYDSSYSYQNT---YSQPSISYQOSS	225
Dc		167	SYS-----QQP-----QQSTVSQQQGGYQCQQGGYCGOSS	197
Qy		226	YQGQSSY-----GQGPSTYSTPYOTGSYSQ-AASQYSQQOSS---YQCP-----SS	266
Dc		198	YSQGGGYSQCTPPQOQOAPPPSSYAPEPSGSGYGPSASQYQGQGSTGAGGASDYKPENQYS	257
Qy		267	FRODDHSMGVYGOBSGGSPGSENRMSGPDNKGRRRGGFDRGMSGRG--RGQ--GRG	323
Dc		258	YRPDHONGGSGYSGPSSGGYGFBBERMGGEENRGRGCFDRGMNGGGCMTGMSRG	317
Qy		324	MSGAGERGFFNRKGGPMDCEGPLDLGPPVDDEDSDNSAIYVGLANDSVTLDDLADFQ	363
Dc		318	MGIAGDRGGSFXKD-----DGEMGAPEED-DSENSTIYITLTENATLLEAVDFKH	369
Qy		384	CGVVKKNKKTQOPMTHLYLDKETGPKCDATVSIJEDPTAKAAEWFDGXDFQSKLKYS	443
Dc		370	SGIIRINKRTGLPAVNIYTDXTGPKCDATLSYEELPPSAKAVAWEFDGXDFQSKLKYS	429
Qy		444	LARKKPRNMNSMKGLPPEBREGMPPLRLGFPGGYGGGGGPGMRG--GRGDGRGGFPPR-G	501
Dc		430	MARRKPRMGMNRMG-----PMKGNDRGMGMRGMGMRGMRGGRGGRGPFMRG	479
Qy		502	PBG-SKGNPSSGGANTQHRAQDMOCNPFCGNQNFAMRTCCNQCAKXPBGFLLPFPBPBG	560
Dc		480	PRGMGNGCPT-CGNMQRAAGDMQCPMNCGNQNFAMMECNQCAKXPBEGGPPPF-PG	537
Qy		561	GDRGRGPGCMRGFRGLMDRGPGGM--FRGHRGDRGFRGGRGMDRGFGGGRRGCP	618
Dc		538	GRRGRGPGCMRGFRG--MDRGPGGPGFRGFRGVRGGR-FRGMDRGGFGG--RGR	591
Qy		619	GGRPGPLMEQMGRRRGGRGGGCKXD-KEHKQERR	652
Dc		592	GGPP--MDWGRRRGRMGPPGCKMDKDRHRR	623
RESULT 13				
Q9BMA2				
ID	Q9BMA2	PRELIMINARY;	PRT:	354 AA.
AC	Q9BMA2			
DT	01-JUN-2001	(TrEMBLrel. 17, Created)		
DT	01-JUN-2001	(TrEMBLrel. 17, Last sequence update)		
DE	01-MAR-2003	(TrEMBLrel. 23, Last annotation update)		
DE		EHSRI protein.		
OS		Homo sapiens (Human).		
OC		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC		Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
OX		[1]	Taxid=9606;	
RN				
RP		SEQUENCE FROM N.A.		
RC		TISSUE=Muscle;		
RX		MEDLINE=42388257; PubMed=12477932; DOI=10.1073/pnas.242603899;		
RA		Strausberg R.L., Feilngold B.A., Grouse L.H., Derge J.G.,		
RA		Klausner R.D., Collins F.S., Wagner L., Shenman C.M., Schuler G.D.,		
RA		Ahtshoch S.P., Zeeberg B., Buotou K.H., Schaefer C.F., Bhat N.K.,		
RA		Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh P.,		
RA		Diatchenko L., Marutina K., Farmer A.A., Rubin G.M., Hong L.,		
RA		Scapleton M., Soares M.B., Bonaldi M.F., Casavant T.L., Scheetz T.E.,		
RA		Bromstein M.J., Usdin T.B., Toshitaki S., Cantinini P., Prange C.,		
RA		Raba S.S., Longueland N.A., Peters G.-J., Abramson R.D., Mullaly S.J.,		
RA		Boesk S.A., McMan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,		
RA		Richards S., Morley K.C., Hale S., Garcia A.M., Gay L.J., Hulik S.W.,		







GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

## OM protein - nucleic search, using frame\_plus.p2n model

Run on: February 20, 2005, 16:33:24 ; Search time 7016.69 Seconds

(without alignments)  
4530.143 Million cell updates/sec

Title: US-10-791-017A-2

Perfect score: 3633  
1 MASTDYSTRSQAAAGCYSA.....GGRGKDKGKHQRDRPY 656

Scoring table:

BLOSUM62	
Xgapop 10.0 , Xgapext 0.5	
Ygapop 10.0 , Ygapext 0.5	
Fgapop 6.0 , Fgapext 7.0	
Delop 6.0 , Delext 7.0	

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0  
Maximum DB seq length: 200000000Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

## Command line parameters:

-MODE=frame+.p2n.model -DEV=xlp  
-O=/cgn2.1/USPTO.spool.p/US10791017/funat.17022005.125807.22074/app.query.fasta\_1.1358  
-DB=GenEmbl -QFMT=faetap -SUFFIX=p2n.rge -MINMATCH=0.1 -HOOPT=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blomsun62 -TRANS=human40.cdi -LIST=45  
-BOCALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=pro -NOM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US10791017@cgn2.1.1.7357@funat.17022005.125807.22074 -NCPU=6 -ICPU=3  
-NO\_WMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

## Database :

GenEmbl:.\*  
1: gb.ba.\*  
2: gb.htg.\*  
3: gb.in.\*  
4: gb.om.\*  
5: gb.ov.\*  
6: gb.pat.\*  
7: gb.ph.\*  
8: gb.pl.\*  
9: gb.pr.\*  
10: gb.ro.\*  
11: gb.sts.\*  
12: gb.sy.\*  
13: gb.un.\*  
14: gb.vl.\*

Prod. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	3633	100.0	1971 12	BT007796
2	3633	100.0	2182 9	BC004817 Homo sapi
3	3633	100.0	2371 6	A36460 Sequence 1
4	3633	100.0	2371 6	AR080100 Sequence

5	3633	100.0	2390 6	CO867364	CO867364 Sequence
6	3633	100.0	2390 6	AX411125	AX411125 Sequence
7	3633	100.0	2390 6	HSEWS	X6689 H. sapiens B
8	3618.5	99.6	2164 9	BC072442	BC072442 Homo sapi
9	3618.5	99.6	2189 9	CR456490	CR456490 Homo sapi
10	3618.5	99.6	2364 9	BC011048	BC011048 Homo sapi
11	3605.5	99.2	2189 9	AK056309	AK056309 Homo sapi
12	3602	99.1	2326 6	CQ730352	CQ730352 Sequence
13	3579	98.5	2189 10	BC068226	BC068226 Mus muscu
14	3571.5	98.3	2188 6	AX305537	AX305537 Sequence
15	3571.5	98.3	2188 10	MMEMS	X79233 M. musculus
16	3564.5	98.1	2396 5	BC083960	BC083960 Xenopus l
17	3291.5	90.6	132906 2	AC109802	AC109802 Canis fam
18	3291.5	90.6	163104 2	AC110672	AC110672 Canis fam
19	3277	90.2	1988 6	AX714390	AX714390 Sequence
20	3277	90.2	1988 6	AK056681	AK056681 Homo sapi
21	3233	89.0	193559 2	AC137500	AC137500 Homo sapi
22	3210	88.4	2026 6	CO850483	CO850483 Sequence
23	3210	88.4	2026 6	AK127624	AK127624 Homo sapi
24	3210	88.4	155815 2	AC018774	AC018774 Homo sapi
25	3210	88.4	180718 9	AL596087	AL596087 Human DNA
26	3210	88.4	182501 2	AC011221	AC011221 Homo sapi
27	3209.5	88.3	2440 5	AJ719366	AJ719366 Gallus ga
28	3156	85.9	221444 2	AC113313	AC113313 Mus muscu
29	3100.5	85.3	247757 2	AC121282	AC121282 Mus muscu
30	2822.5	77.7	1783 6	CQ721057	CQ721057 Sequence
31	2819.5	77.6	226400 2	AC128482	AC128482 Rattus no
32	2819.5	77.6	235241 2	AC098231	AC098231 Rattus no
33	2819.5	77.6	263925 2	AC106522	AC106522 Rattus no
34	2730	75.1	7293 9	HSMB08920	BSX64872 Homo sapi
35	2601	71.6	2665 5	BC063928	BC063928 Xenopus t
36	2586.5	71.2	2683 5	BC075120	BC075120 Xenopus t
37	2338	64.4	270337 2	AC099212	AC099212 Rattus no
38	2292	63.1	221924 2	AC105889	AC105889 Rattus no
39	2292	63.1	251050 2	AC118907	AC118907 Rattus no
40	1908.5	52.5	211189 2	AC128031	AC128031 Rattus no
41	1908.5	52.5	217080 10	AC125565	AC125565 Rattus no
42	1885.5	51.9	2489 5	BC068357	BC068357 Dario rer
43	1880	51.7	2448 5	BC056281	BC056281 Dario rer
44	1816	50.0	2473 9	AK026270	AK026270 Homo sapi
45	1786	49.2	1864 9	AF254086	AF254086 Homo sapi

## ALIGNMENTS

RESULT 1  
LOCUS BT007796 1971 bp mRNA linear SYN 13-MAY-2003  
DEFINITION Synthetic construct Homo sapiens Ewing sarcoma breakpoint region 1  
ACCESSION BT007796  
VERSION BT007796.1 GI:30584430  
KEYWORDS FLI CDNA.  
SOURCE synthetic construct  
ORGANISM other sequences; artificial sequences.  
REFERENCE 1 (bases 1 to 1971)  
Kaulingya,N., Chen,X., Rolfs,A., Halleck,A., Hines,L., Eisenstein,S.,  
Kaulingya,M., Raphael,J., Moreira,D., Kelley,T., Labaer,J., Lin,Y.,  
Pheasant,M. and Farmer,A.  
Cloning of human full-length cDNAs in BD Creator(TM) System Donor  
vector  
JOURNAL Unpublished  
2 (bases 1 to 1971)  
Kaulingya,N., Chen,X., Rolfs,A., Halleck,A., Hines,L., Eisenstein,S.,  
Kaulingya,M., Raphael,J., Moreira,D., Kelley,T., Labaer,J., Lin,Y.,  
Pheasant,M. and Farmer,A.  
TITLE Direct Submission  
SUBMITTED (13-MAY-2003) BD Biosciences Clontech, 1020 East Meadow  
Circle, Palo Alto, CA 94303, USA  
COMMENT This cDNA clone is a part of a collection of human full length  
expression clones generated by BD Biosciences Clontech and the  
Harvard Institute of Proteomics. Each cDNA has been cloned in two





Db 20 ATGGCGTCCAGGATTAACAGTACCTTAAGCCAGCTGACGAGCGGCTACAGTGGT 79  
 QY 21 TTTTThAlAGlnProThrGlnGlyTyrAlaGlnThrThGlnAlaTyrGlnGlnSer 40  
 Db 80 TACACGCCGACGCTCAAGATATGACACACACCGACCATATGGCCAAAC 139  
 QY 41 TTTGThTyrGlnProThrAspValSerTyrThGlnAlaGlnThrThAlaThr 60  
 Db 140 TATGAACTATGAGAACGCCCATGATGCTATACCCAGGCTCAAGCACTGAC 199  
 QY 61 TTTGThAlaThAlaTyrAlaThrSerTyrGlnProProThrGlnTyrThThrPro 80  
 Db 200 TATGGGACAGCCGCTATGCACTTCTATGACAGCTCCCACTGGTATATCTCA 259  
 QY 81 ThAlaProGlnAlaTyrSerGlnProValGlnGlyTyrGlnTyrAlaTyrAspThr 100  
 Db 260 ACTGCCCCCAGGACAAACAGCCAGCTGTCAGGGGTATGGCACTGGTCTATGATACC 319  
 QY 101 ThThAlaThAlaTyrThThThGlnAlaSerTyrAlaAlaGlnSerAlaTyrGlnTyr 120  
 Db 320 ACCACTGCTACAGTCAACACACCGGCTCTTATGAGCTCAAGTCTGATATGGCACT 379  
 QY 121 GlnProAlaTyrProAlaTyrGlnGlnProAlaAlaThAlaProThrArgProGln 140  
 Db 380 CAGCTGCTTATCCAGCTATGGGACAGCCAGCCAGCTGCACTCAACAGACCGCAG 439  
 QY 141 AspGlyAsnIleProThrGlnTyrThSerGlnProGlnSerThrGlnTyrAlaGln 160  
 Db 440 GATGGAACAAAGCCCATGAGTACATCACTCAATCTAGCACAGGGGTATCAACAG 499  
 QY 161 ProSerLeuGlyTyrGlnGlnSerAsnTyrSerTyrProGlnAlaProGlnSerTyrPro 180  
 Db 500 CCAAGCTAGAGATATGACAGATTAACATCACTTATCCCAAGTACTGGAGCTACCC 559  
 QY 181 MetGlnProValThAlaProProSerTyrProProThrSerTyrSerSerThrGlnPro 200  
 Db 560 ATGCAACAGTCACTGACCTCATCTTACCTTACCAAGCTATCTCTACAGCCG 619  
 QY 201 ThSerTyrAspGlnSerSerTyrSerGlnGlnAlaThThTyrGlnProSerSerTyr 220  
 Db 620 ACTAGTATGATACAGAGTATCTCTCAGACAAACCTATGGGCAACCGACAGCTAT 679  
 QY 221 GlnGlnGlnSerSerTyrGlnGlnGlnSerSerTyrGlnGlnProProThrSerTyr 240  
 Db 680 GGAACACAGATGCTATGCTCAACAGACGCTATGGGACAGCTCCCACTAATTAC 739  
 QY 241 ProProGlnThrGlySerTyrSerGlnAlaProSerGlnTyrSerGlnGlnSerSer 260  
 Db 740 CCAACCCAACTGATCTTACAGCCAAAGCTCAAGTCAATATAGCCAAACAGACAGCAGC 799  
 QY 261 TTTGThGlnGlnSerSerPheArgGlnAspPheProSerSerMetGlyValTyrGln 280  
 Db 800 TACGGGACAGAGTTCATTCACAGACACCCAGTAGAGTGGTGTATTTGGGACG 859  
 QY 281 GluSerGlyTyrPheSerGlyProGlnGlnAsnArgSerMetSerGlyProAspAsnArg 300  
 Db 860 GAGTCTGAGAGATTTCCGACCAAGAGAGAACCCGAGCATGAGTGGCTGTATACCGG 919  
 QY 301 GlyArgGlyValArgGlyPheAspArgGlyGlyMetSerArgGlyGlyArgGlyGly 320  
 Db 920 GGCAGGGGAAAGGGGATTTGATCGTGAGGATGACAGAGTGGGCGGGAGAGAGG 979  
 QY 321 ArgGlyGlyMetGlySerAlaGlyGlnArgGlyGlyPheAsnIleProGlnGlyPromet 340  
 Db 980 CGGGTGGAAATGGGACGGCTGAGAGCGAGTGGCTTCAATAGCCCTGTGTGAGCCATG 1039  
 QY 341 AspGlnGlyProAspLeuAspLeuGlyProProValAspProAspGlnAspSerAspAsn 360  
 Db 1040 GATGAAGAGACCAATCTTATCTAGGCCCACTGTAGATCAATGAAGATCTTGACAC 1099  
 QY 361 SerAlaIleTyrValGlnGlyLeuAsnAspSerValThrLeuAspLeuAlaAspPhe 380  
 Db 1100 AGTGAATTTATGATACAGAGATTAATGACAGTGTGCTTATGATGATCTGCGAGACTTC 1159

QY 381 PheArgGlnCysGlyValValIleMetAsnIleArgThrGlnProMetIleHisIle 400  
 Db 1160 TTTAAGCAGTGTGGGTTGTATGAATGAACAAAGAACTGGGCAACCCATGATCCATC 1219  
 QY 401 TTTLeuAspIleGlyThrGlyLysProIleGlyAspAlaThrValSerTyrGlnAspPro 420  
 Db 1220 TACCTGACAAAGAAACAGAAAGCCCAAGGCGATGCGACAGTGTCTTATGAAGCCCA 1279  
 QY 421 ProThrAlaValAlaValAlaIleTyrPheAspGlyLysAspPheGlnGlySerLysLeu 440  
 Db 1280 CCCACTGCGCAAGCTCGCGTGGAAATGTTTATGAGAAAGATTTTCAGGGAGCAAACT 1339  
 QY 441 LysValSerLeuAlaArgLysIleAspProPheMetAsnSerMetArgGlyLysLeuProPro 460  
 Db 1340 AAAGTCTCTTGTCTGGAAGAGCTCCAAATGAACATGATCGGGGTGTCTGCCACCC 1399  
 QY 461 ArgGlnGlyArgGlyMetProProProLeuArgGlyGlyProGlnGlyProGlnGlyPro 480  
 Db 1400 CGTGAGGGGAGAGGATGCAACCACTCCGTGAGAGTCCAGAGGCCCAAGAGGTCT 1459  
 QY 481 GlyIlePrometGlyArgMetGlyLysArgGlyLysAspArgGlyLysPheProProArg 500  
 Db 1460 GGGGAGCCCATGGTGTGATGGAGAGCGCTGAGAGAGATGAGAGAGGCTTCCCTCCACA 1519  
 QY 501 GlyProArgGlySerArgGlyValAsnProSerGlyGlyLysAsnValGlnHisArgAlaGly 520  
 Db 1520 GAACTCCGGGGTCCCGAGAGAACTCTGAGAGAGAAAGATCTCAGACCGAGCTGGA 1579  
 QY 521 AspTyrGlnCysProAsnProGlyCysGlyAsnGlnAsnPheAlaTyrArgThrGlnCys 540  
 Db 1580 GACTGGCAGTGTCCCAATCCGGGTTGTGAAACAGAACTTCCGCTGAGAAACAGAGTGC 1639  
 QY 541 AsnGlnCysValAlaProLysProGlnGlyPheLeuProProProPheProProGly 560  
 Db 1640 AACCAATGTAGAGCCCAAGCTGAAGCTTCTCCGCGCACCTTTCCGCGCGGGT 1699  
 QY 561 GlyAspArgGlyArgGlyGlyProGlnGlyMetArgGlyGlyArgGlyGlyLysLeuMetAsp 580  
 Db 1700 GGTATCTGGCAGAGTGGCCCTGTGTGATCGGGAGAGAGAGAGTGGCTCATGAT 1759  
 QY 581 ArgGlyGlyProGlnGlyMetPheArgGlyGlyArgGlyGlyAspArgGlyLysPheArg 600  
 Db 1760 CGTGGTGTCCCGGTGAATGTCAAGGTGCGCGGTGAGAGACAGAGTGGCTTCCGT 1819  
 QY 601 GlyIleValArgGlyMetAspArgGlyLysPheGlyGlyLysArgGlyGlyProGlnGly 620  
 Db 1820 GGTGGCCGGGCGATGACCGAGTGTCTTGTGTGAGAGAGACAGAGTGGCTTGGGGGG 1879  
 QY 621 ProProGlnProLeuMetGlnGlnMetGlyGlyArgArgGlyGlyArgGlyGlyProGln 640  
 Db 1880 CCCCTGACCTTGTATGAGACAGATGGAGAGAGAGAGAGAGAGCTGAGAGACTTGA 1939  
 QY 641 LysMetAspIleGlyLysIleArgGlnGlnArgArgAspArgProTyr 656  
 Db 1940 AAAATGATAAAGGAGAGACCGTCAAGAGGCCCAAGATGGCCCTTAC 1987

RESULT 3  
 A36460  
 LOCUS 2371 bp DNA linear PAT 05-MAR-1997  
 DEFINITION Sequence 1 from Patent WO9323549.  
 ACCESSION A36460  
 VERSION A36460.1 GI:2293778  
 KEYWORDS  
 ORGANISM Homo sapiens (human)  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 1 (bases 1 to 2371)  
 AUTHORS Auriat,A., Delattre,O., Desmaze,C., Melot,T., Peter,M.,  
 Plougaestel,B., Thomas,G. and Zucman,D.  
 NUCLEIC ACID CORRESPONDING TO A GENE OF CHROMOSOME 22 INVOLVED IN  
 RECURRENT CHROMOSOMAL TRANSLOCATIONS ASSOCIATED WITH THE



DEVELOPMENT OF CANCEROUS TUMORS  
Patent: WO 9323549-A.1 25-NOV-1993;

CENTRE NAT RECH SCIENT (FR)

Other publication FR 2691475 931126

Other publication JP 850964T 960206.

Location/Qualifiers

1. .2371

/organism="Homo sapiens"

/mol\_type="unassigned DNA"

/db\_xref="taxon:9606"

/dev\_stage="FOETUS"

# ORIGIN

## Alignment Scores:

Pred. No.:	6.22e-77	Length:	2371
Score:	3633.00	Matches:	656
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	6	Gaps:	0

US-10-791-017a-2 (1-656) x A36460 (1-2371)

```

QY      1 MetAlSeThrAspTySerThrTySerGlnAlaAlaGlnGlnGlnTySerAla 20
DB      25 ATGGCTCCAGGATTAACGACTTATAGCCAGCTGCGAGCGAGCGGCTACAGTGGCT 84
QY      21 TyrThAlaGlnProThrGlnGlyTyrAlaGlnThrGlnAlaTyrGlyGlnGlnSer 40
DB      85 TACACGGCCAGCCCACTGAGGATATGACACACCGGCAATATGGCGAACAACAGC 144
QY      41 TyrGlyThrTyGlyGlnProThrAspValSerTyrThrGlnAlaGlnThrAlaThr 60
DB      145 TATGGAACCTATGAGACAGCCCACTGATGCAAGCTATACCCAGGCTCAGACCACTGCAACC 204
QY      61 TyrGlyGlnThrAlaTyrAlaThrSerTyrGlyGlnProProThrGlyTyrThrThrPro 80
DB      205 TATGGGCAACCCGCTTATGCACTTCTTATGACAGCTCCCACTGATTAATCTACTCCA 264
QY      81 ThrAlaProGlnAlaTyrSerGlnProValGlnGlyTyrGlyThrGlyAlaTyrAspThr 100
DB      265 ACTGCCCCCAAGCATACAGCCAGCTGCTCAGAGGATATGGCGACTGGTCTTATGATAC 324
QY      101 ThrThrAlaThrValThrThrThrGlnAlaSerTyrAlaAlaGlnSerAlaTyrGlyThr 120
DB      325 ACCACGCTACAGTGCACCAACCCAGGCTCTCTATGAGCTCAGTCTGATATGCACT 384
QY      121 GlnProAlaTyrProAlaTyrGlyGlnGlnProAlaAlaThrAlaProThrThrProGln 140
DB      385 CAGCTGCTTATTCACCTATGAGGCGAGCGAGCGACCACTGCACTTACAGACCGCAG 444
QY      141 AspGlyAsnLysProThrGlnThrSerGlnProGlnSerSerThrGlyTyrAsnGln 160
DB      445 GATGGAAACAGGCCCACTGAGTACTGACACTCAATCTAAGCAAGGGGTTTACACCGAG 504
QY      161 ProSerLeuGlyTyrGlyGlnSerAsnTyrSerTyrProGlnValProGlySerTyrPro 180
DB      505 CCCAGGCTGAGATGAGACAGAGTACTCAGATTATCCCAAGTACCTGGAGCTACCCCC 564
QY      181 MetGlnProValThrAlaProProSerTyrProProThrSerTyrSerSerThrGlnPro 200
DB      565 ATGCAACCACTACCTGCACTCATCTCACTTACCTTACCAAGCTATCTTACACAGCGG 624
QY      201 ThrSerTyrAspGlnSerSerTyrSerGlnGlnAlaAsnThrTyrGlyGlnProSerSerTyr 220
DB      625 ACTAGTTATGATCAGACGAGTAACTCTCAGCAGAAACCTATGGCGAACGAGCACTAT 684
QY      221 GlyGlnGlnSerSerTyrGlyGlnGlnSerSerTyrGlyGlnGlnProProThrSerTyr 240
DB      685 GGAAGCAAGATAGCTATGCTCAACAAACAGCTATAGGCGAGAGCTCCCACTAGTTAC 744
QY      241 ProProGlnThrGlySerTyrSerGlnAlaProSerGlnTyrSerGlnGlnSerSerSer 260

```

```

DB      745 CCAACCCCAACTGCATCTTACAGCCAGCTCCAAGTCAATATATGCCAACAGACGACGAC 804
QY      261 TyrGlyGlnGlnSerSerPheArgGlnAspHisProSerSerMetGlyValTyrGlyGln 280
DB      805 TACGGGCAACAGAGTTCAATCCGACAGACACCCCAAGTACATGGGTGTTATGGCGAG 864
QY      281 GluSerGlyGlyPheSerGlyProGlyGlyLysAsnLysSerMetSerGlyProAspAsnArg 300
DB      865 GAGTCTGAGAGATTTTCCGACCCAGAGAAACCCGAGAGATGATGGCTTGATTAACCCG 924
QY      301 GlyArgGlyArgGlyValPheAspArgGlyGlyMetSerArgGlyValArgGlyGlyGly 320
DB      925 GCGAGGGAAGAGGGGATTTGATCTGAGGCAATAGCAGAGTGGCGGGAGAGAGA 984
QY      321 ArgGlyGlyMetGlySerAlaGlyGlnArgGlyGlyPheAsnLysProGlyGlyProMet 340
DB      985 CGCGGTGAGATGGGACGCGCTGGAGAGCGAGGTGCTTCAATAGCTGCTGGTGAACCATG 1044
QY      341 AspGlnGlyProAspLeuAspLeuGlyProProValAspProAspGlyLysAspAsn 360
DB      1045 GATGAAGACACAGATCTTATCTAGGCTCTCTGATTCAGATGAAAGACTTGAACAC 1104
QY      361 SerAlaIleTyrValGlnGlyLeuAsnAspSerValThrLeuAspAspLeuAlaAspPhe 380
DB      1105 AGTCAATTTATGTCACAGATTTAATGACAGTGTACCTTACATGATCTGGCAGACTTC 1164
QY      381 PheLysGlnCysGlyValValLysMetAsnLysArgThrGlyGlnProMetIleHisIle 400
DB      1165 TTTAAGCAGATGATGGGTGTTTAAAGATGACAAAGAACTGGGCAACCATGATCCACATC 1224
QY      401 TyrLeuAspLysGlnThrGlyLysProLysGlyAspAlaThrValSerTyrGluAspPro 420
DB      1225 TACCTGGACAAAGAAACAGAAAGCCCAAGGCGAAGCCACAGTGTCTTATGAAAGCCA 1284
QY      421 ProThrAlaLysAlaAlaValGlnThrPheAspGlyLysAspPheGlnGlySerLysLeu 440
DB      1285 CCACGCTCCAAAGGCTGCTGGAATGTTGATGGAAAGATTTTCAAGGAGCAAACTT 1344
QY      441 LysValSerLeuAlaArgLysLysProProMetAsnSerMetLysGlyLysLeuProPro 460
DB      1345 AAAGTCTCCCTTGTCTGGAAAGAGCTTCAATGAAACAGATATGCGGGGTGCTGCCACCC 1404
QY      461 ArgGlnGlyArgGlyLysProProProLeuAlaGlyGlyLysProGlyLysProGlyLysPro 480
DB      1405 CGTGAAGGAGAGGAGATGCAACACCACTCCGTGAGGTCACAGAGGCCACAGAGGTCT 1464
QY      481 GlyGlyProMetGlyArgMetGlyGlyArgGlyGlyLysAspArgGlyGlyLysPheProProArg 500
DB      1465 GGGGAGCCATGGGTGCTATGGAGAGCCGTGAGAGGATAGAGAGGCTTCCCTCCAAAG 1524
QY      501 GlyProArgGlySerArgGlyLysAsnProSerGlyGlyLysAsnValGlnHisArgAlaGly 520
DB      1525 GGAACCCCGGGGTTCCGAGAGAAACCTCTGAGAGGAGAAAGCTCCAGACACGAGCTGGA 1584
QY      521 AspTyrGlnCysProAsnProGlyCysGlyLysAsnGlnAsnPheAlaThrArgThrGlnCys 540
DB      1585 GACTGCAAGTGTCCCAATCCGGGTGTGGAACCAAGAACTTCCCTCGGAGAAACAGAGTGC 1644
QY      541 AsnGlnCysLysAlaProLysProGlnGlyPheLysProProProPheProProProGly 560
DB      1645 AACCAATGTAAGCCCAAGACCTGAAAGCTTCTCCGCAACCTTTCGCGCCCGGAGT 1704
QY      561 GlyAspArgGlyLysGlyLysProGlyGlyMetArgGlyGlyLysGlyLysLeuMetAsp 580
DB      1705 GGTGATCGGAGAGGTGCGCTGATGCAAGCGGAGGAGAAAGAGTGGCTCATATGAT 1764
QY      581 ArgGlyGlyProGlyGlyLysMetPheArgGlyGlyLysArgGlyGlyLysPheArg 600
DB      1765 CGTGTGTGTCCCGGAGATGTTCAGAGGTGCGGTGTGAGAGACAGAGTGGCTTCCCT 1824
QY      601 GlyGlyLysGlyLysMetAspArgGlyGlyLysPheGlyGlyGlyLysArgArgGlyLysProGly 620
DB      1825 GGTGGCCGGGAGATGACCGAGGTGCTTGTGTGAGAGAAAGCGAGTGGCTTGGGGGG 1884

```



Oy		521	AAPTTPGSLCYAPROAANPProGlyLVYveGLYAAnGIlaanPIhealATPraxThnGluLys	540
Dd		1585	GACTGGCAGTGTCCCAATTCGGGGTTGTGAACCAAACTTGCGCTGGAGAACAAGAGTGC	1644
Oy		541	AenGIcYtLYBALAPProLYABProGUlYLPhelaUPProProProPhoProProGly	560
Dd		1645	AACCAGTATTAAAGGCCCAAAGCCTTAAGGCTTTCCTCCGCCACCCTTTCCGGCCCCGGGT	1704
Oy		561	GLYABPAARGGLYARGLLYGLYPProGLYGLYMecArGLLYGLYASGLLYGLYLEumecAp	580
Dd		1705	GGTGATCTCTGGGAGAGGTGGCCCTTGCTGTGCATGCGGGGAGGAAGAAGGTGGCTCATGTGAT	1764
Oy		581	ARGGLYGLYProGLYGLYMecPhearGLLYGLYARGLLYGLYABPAARGGLLYGLYPhearG	600
Dd		1765	CCTGGTGGTCCCCTGGTAGAATGTTCAAGAGGTGGCCCTGGTGGAGACAGAGTGGCTTCCGT	1822
Oy		601	GLYGLYARGLLYMecABPAARGGLLYGLYPheGLYGLYGLYARGARGLLYGLYPProGLYGLY	620
Dd		1825	GGTGGCCGGGGGATGATGACCAAGGTGGCTTTGGTGGAGGAAGACAGAGTGGCCCTGGGGGG	1886
Oy		621	PROProGLYProLEumEteGLUGImetGLYGLYARGARGLLYGLYARGLLYGLYPProGLY	640
Dd		1885	CCCCCTGGACCTTTTATGTGAAACAGATGGGAGGAAGAAGAGGAGCAGTGGAGSACCTGGA	1944
Oy		641	LYMeTABRPLYGGLYLUHLaRGLLInGLIUAArgAARGAPARGProTYr	656
Dd		1945	AAAATGATTAAGGCGACACCGTGAAGAGCGCAAGATCGGCTCAC	1992
RESULT 5				
LOCUS	COB67364	2390 bp	DNA	linear PAT 13-SEP-2004
DEFINITION	Sequence 1 from Patent EP1455190.			
ACCESSION	COB67364			
VERSION	COB67364.1	GI:5197589		
SOURCE	Homo sapiens (human)			
ORGANISM	Homo sapiens			
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
	Mammalia; Euteria; Primates; Catarrhini; Hominiidae; Homo.			
REFERENCE	Obendorf,M. and Wolf,S.			
AUTHORS	Modulation of the activity of nuclear receptors via EMS			
TITLE	Patent: EP 1455190-A 1 08-SEP-2004;			
JOURNAL	Schering Aktiensgesellschaft (BE)			
FEATURES	Location/Qualifiers			
source	1..2390			
	/organism="Homo sapiens"			
	/mol_type="unassigned DNA"			
	/db_xref="taxon:9606"			
	44..2014			
	/note="unnamed protein product"			
	/codon_start=1			
	/protein_id="CAH33891.1"			
	/db_xref="GI:5197590"			
	/translation="MASPDVSTYSQAAMQGSAYTAQPNOGYAQTTOAVGSGSYGVTV			
	GPDPDVSTQAQTATTTGCTATVATSYGCPPTGTTTTPTAPKAYSQPVQGYGTGADITTT			
	ATVTTOASVIAQSAVGTQAPVPAYAGQDPAATAPTRPDGNKPETTSOPSSGTGINP			
	PSLGQGQENSVYQQVGSYPMQPVTPAPSPYEPTSSYPTSIDSYSTOONTYGQEP			
	SYGOOSISYGOOSISYQCPPTSPQPTGYSQALPSYSGOOSISYGOOSIFRDHDHSMQM			
	VYGGSSGFGFSGEGENRSMMSGPDNRRGRGGFDRCMSGGRGSGAGRGSGAGEGSRG			
	KRGMRDBRPLDLGPPVDPEEDNSAIIVYGLINDSYTLDDLADFPGOCVVIMNKIK			
	TGQPMITHILYDKETSKPKGDATVSIEDPTAKAAYEMFRDGDFQSGKLVSILARKKRP			
	NWSNMKGLPFRERGRMPPLRLRGPGPGPGGGPMWRMGSGDRGFPRPSRSRKRP			
	PSGGCNVQHRAGDMQCNPNGCSNONFAWTBCNOCAPKPSGLPAPP PGCDGRGRGR			
	GPSGRTGCRGGLMDRGPCGMFRGCRGDRGGRGMDRGFGRRGMDRGFGGGRGCGPGRPG			
	LMEQNGRRGGRGCGPRKMDKGEHRRERDRPY"			
ORIGIN				
Alignment Scores:	6,25e-77	Length:	2390	
Pred. No.:	3633.00	Matches:	656	
Percent Similarity:	100.00%	Conservative:	0	

DB	Query Match:	Best Local Similarity:	Mismatches:	Gaps:
DB: 100.00%	100.00%	0	0	0
US-10-791-017A-2 (1-656) x C0867364 (1-2390)				
QY	1 MetalaseThraapTySerThThTySerGlnAlaAlaGlnGlnGlyTySerAla	20		
DB	44 ATGGGCTTCAGAGATTACAGTACCTATAGCCAAAGCTGAGGGCAGAGGGCTACAGTCT	103		
QY	21 TyThrAlaGlnProThrGlnGlyTyAlaGlnThThrGlnAlaTyGlyGlnGlnSer	40		
DB	104 TACACCGCCAGCCCACTCAAGATATGCAAGACACCCAGCATATGGGCACAAAGC	163		
QY	41 TyGlyThThTyGlyGlnProThraAspValSerTyThrGlnAlaGlnThThraAlaTh	60		
DB	164 TATGGAACTATWGACAGCCCACTGATGTGAGTATACCAGGCTCAGACCACTGCAACC	223		
QY	61 TyGlyGlnThThraAlaTyAlaThSerTyGlyGlnProProThrGlnTyThThThPro	80		
DB	224 TATGGGCAAGCCGCTATGACAATCTTATWGACAGCCTCCACTGGTTATATCACTCA	283		
QY	81 ThraAlaProGlnAlaTySerSerGlnProValGlnGlyTyGlyThThraAlaTyAspThr	100		
DB	284 ACTGCCCCCAAGGCATACAGCAGCTCTCCAGGGATATGGCACTGGTGCTTATGATACC	343		
QY	101 ThThraAlaThraValThThThThrGlnAlaSerTyAlaAlaGlnSerAlaTyGlyThr	120		
DB	344 AACAAGCTGATACATGACCAACCAAGGCTCTCTATGAGCTCATGTGATATGGCACT	403		
QY	121 GlnProAlaTyProAlaTyGlyGlnGlnProAlaAlaThraAlaProThraGProGln	140		
DB	404 CAGCTGCTTATCCAGCCTATGGCAGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG	463		
QY	141 AspGlyAsnTyProThrGlnThSerGlnProGlnSerSerThGlnGlyTyAspGln	160		
DB	464 GATGAAACAAGCCCACTGAGACTAGTACCTCAATCTTACACAGGGGGTTACACAGC	523		
QY	161 ProSerLeuGlyTyGlyGlnSerAsnTySerTyProGlnValProGlySerTyPro	180		
DB	524 CCGAGCTTNGATATGACAGATGATCTACAGTTATCCCAAGTATCCTGGAGCTACCC	583		
QY	181 MetGlnProAlaThraAlaProProSerTyProProThrSerTySerSerThGlnPro	200		
DB	584 ATGAGCAGCAGTCACTGACCTCCATCTTACCTCTTACCACTATTCCTTACACAGCG	643		
QY	201 ThSerTyThraAspGlnSerSerTySerGlnGlnAsnThTyGlyGlnProSerSerTy	220		
DB	644 ACTAGTATATGATCAGAGAGATTACTCTCAGAGAAACCTATGGGCAACCCAGAGCTAT	703		
QY	221 GlyGlnGlnSerSerTyGlyGlnGlnSerSerTyGlyGlnGlnProProThSerTy	240		
DB	704 GAGACGACAGATGCTATGTGACAAAGACGCTATGGCAGAGCTTCCACTAGATTAC	763		
QY	241 ProProGlnThThTySerTySerGlnAlaProSerGlnTySerGlnGlnSerSerSer	260		
DB	764 CCACCCCAAATGATCTCTACAGCAGACTCCAAGTCAATATACCAAGAGCAGCAGC	823		
QY	261 TyGlyGlnGlnSerSerPheArgGlnAspHisProSerSerMetGlyValTyGlyGln	280		
DB	824 TACGGGCGACGAGATTCACTCCGACAGAGACCAACCCAGTACAGGGGTATTATGGCAG	883		
QY	281 GluSerGlyGlyPheSerGlyProGlyGlnAsnArgSerMetSerGlyProAspAsnArg	300		
DB	884 GAGCTTGAGGAGATTTTCCGACACAGAGAGAAACCGAGACAGAGTGGCTTGATACCGG	943		
QY	301 GlyATGAGlyATGAGlyGlyPheAspArgGlyGlyMetSerATGAGlyGlyArgGlyGly	320		
DB	944 GCGAGGGGAAGAGGGGATTTGATCTGTGAAGCATGACAGAGTGGCGGGGAGAGAGGA	1003		
QY	321 ArgGlyGlyMetGlySerAlaGlyGlyLysArgGlyGlyPheAsnTyProGlyGlyProMet	340		
DB	1004 CCGGGTGGAAATGGGAGCGCTGAGAGAGGAGGTGGCTTCATTATAGCCTGGTGAACCATG	1063		

QY 341 AARGGLUGLYPROABSLUENGLYPROPROVALASPPOAGLUAAPSERAPAPAN 360  
 DB 1064 GATGAAGAGACCAATCTTGATCTAGGCCACCTGATGATCAATAGAGCTCTGACAC 1123  
 QY 361 SerAla1LeuValGlnGlyLeuAanAapSerValThreLeuAapLeuAlaAepPie 380  
 DB 1124 AGTGAATTTATGTAACAGATTTAAATGACAGTGTACCTGATGATGATGAGACTTC 1183  
 QY 381 PheLeuGlnCysGlyValValValValMetAanLysArgThyGlnProMetIleHISle 400  
 DB 1184 TTTAAACAGTGTGGGTGTTTAAAGATGACAAAGAACTGGGCAACCATGATCCACATC 1243  
 QY 401 TyrLeuAapLysGlyLeuThreGlyLysProLysGlyAAspAlaThrValSerTyrgLUAAPPro 420  
 DB 1244 TACCTGACAAAGAAACAGAAAGCCCAAGGCAAGCCACAGTGTCTATGAAAGACCA 1303  
 QY 421 ProThra1AlaValAlaValAlaValTTPheAapGlyLysAapPheGlnGlySerLysLeu 440  
 DB 1304 CCACACTGCCAAGGCTCCGTGGAATGGTTGATGGGAAAGATTTCAAGGAGCAAACTT 1363  
 QY 441 LysValSerLeuAlaArgLysLysProProMetAanSerMetArgGlyLysLeuProPro 460  
 DB 1364 AAAGTCTCCCTGCTCGAAGAAAGCTCCATGATGATGCGGGTGTCTGCAACCC 1423  
 QY 461 AARGGLUGLYAARGGLYMETPROPROPLOLEUAARGGLYLYPROGLYGLYLYPRO 480  
 DB 1424 CGTGAAGGAGAGAGCAATGCAACCACTCTCGTGAAGTTCAGAGAGGCCCAAGAGGTCT 1483  
 QY 481 GLYGLYPROMETGLYARGMETGLYGLYARGGLYGLYLYAAspArgGlyLysPheProProArg 500  
 DB 1484 GGGGAGCCATGTGTGTCATGGAGGCGGTGAGAGATGAGAGAGGCTTCCCTCAAG 1543  
 QY 501 GLYPROARGLYSERARGLYAANPROSERGLYGLYLYAAsnValGlnHISArgAlaGly 520  
 DB 1544 GGACCCCGGGGTTCAGAGAGCAACCTCTGAGAGAGAAACCTCCAGCAACCAAGCTGA 1603  
 QY 521 APTTGTGLCYAPROAANPROGLYCYBGLYAAsnGlnAanPheAlaThrArgThraGlyCys 540  
 DB 1604 GATGCGAATGTCCTCAATCCGAGGTGTGAAACCAAGAACTTCCCTGAGAAACAGATGC 1663  
 QY 541 AAsnGlnCysLysAlaProLysProGlnGlyPheLeuProProProPheProProGly 560  
 DB 1664 AACCAATGTAAGGCCCAAGCTGAAAGCTTCTCCGCACTTCCGCGCCCGGG 1723  
 QY 561 GLYAspArgGlyAARGGLYGLYLYPROGLYGLYLYMETARGGLYGLYLYLEUWETARP 580  
 DB 1724 GGTGATCTGGCAGAGGTGGCCCTGTGTGTCATCGGGAGAGAAAGAGTGGCTCATGAT 1783  
 QY 581 AARGGLYGLYPROGLYGLYMETPHEARGGLYGLYARGGLYGLYLYAAspArgGlyLysPheArg 600  
 DB 1784 CGTGTGTGTCTCCGATGATTTTCAAGATGTGTCGTGTGTGAGACAGAGGTGTCTTCCT 1843  
 QY 601 GLYGLYARGGLYMETAspArgGlyGLYPheGlyGLYGLYLYARGARGGLYGLYLYPROGLYGLY 620  
 DB 1844 GGTGTGTGTCTCCGATGATTTTCAAGATGTGTCGTGTGTGAGACAGAGGTGTCTTCCT 1903  
 QY 621 PROPROGLYLYPROLEUWETGLYGLYGLYLYMETGLYGLYLYARGARGGLYGLYLYLYPROGLY 640  
 DB 1904 CCCCCCTGAGACTTTGATGAAACAGATGGAGAGAAAGAGAGAGAGAGAGAGAGAGAGAGAG 1963  
 QY 641 LysMetAspLysGlyLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 656  
 DB 1964 AAAATGATTAAGGCGAGACACCTTCAGAGAGCGCAAGATGCGCCCTAC 2011

RESULT 6  
 AX411125  
 LOCUS AX411125 2390 bp DNA 1 linear PAT 14-JUN-2002  
 DEFINITION Sequence 3772 from Patent WO0229103.  
 ACCESSION AX411125 GI:21443830  
 VERSION AX411125.1  
 KEYWORDS Homo sapiens (human)  
 SOURCE

ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE  
 1 Alvarez, C., Horne, D., Peres-da-Silva, S. and Vockley, J. G.  
 TITLE Gene expression profiles in liver cancer  
 JOURNAL Patent: WO 0229103-A 3772 11-APR-2002;  
 GENE LOGIC INC (US)  
 FEATURES  
 source location/Qualifiers  
 1..2390  
 /organism="Homo sapiens"  
 /mol\_type="unassigned DNA"  
 /db\_xref="taxon:9606"  
 /note="EMBL/GenBank Accession No. X66899"

ORIGIN  
 Alignment Scores:  
 Pred. No.: 6,25e-77 Length: 2390  
 Score: 3633.00 Matches: 656  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: 6 Gaps: 0

US-10-791-017A-2 (1-656) x AX411125 (1-2390)

QY 1 MetAlaSerThrAspTyrSerThrTyrSerGlnAlaAlaAlaGlnGlnGlyTyrSerAla 20  
 DB 44 ATGGGCTCCAGGATTAACAGTACCTATAGCCAGGCGCAGAGGAGGCTTACAGTGTCT 103  
 QY 21 TyrThrAlaGlnProThreGlnGlyTyrAlaGlnThrGlnAlaTyrGlyGlnGlnSer 40  
 DB 104 TACACCGCCCAACCCCACTCAAGATATGACAGACCAACCAAGGATATGGGCAACAAGC 163  
 QY 41 TyrGlyThrTyrGlyGlnProThraPValSerTyrThrGlnAlaGlnThrThraLthr 60  
 DB 164 TATGGAACCTATGACAGCCCACTGATGTCACTATACCAAGGCTCAGACCACTGCACCC 223  
 QY 61 TyrGlyGlnThrAlaTyrAlaThrSerTyrGlyGlnProProThraGlyTyrThrThrPro 80  
 DB 224 TATGGCAGACCGCCTATGCAACTTATGAGACGCTCCACATGTTACTACTCA 283  
 QY 81 ThrAlaProGlnAlaTyrSerGlnProValGlnGlyTyrGlyTyrGlyLysAlaThr 100  
 DB 284 ACTGCCCCCAAGCATACAGCAAGCTGTCCAGGGATGAGCACTGTCTTATGATAC 343  
 QY 101 ThrThrAlaThrValThrThrThrGlnAlaSerTyrAlaAlaGlnSerAlaTyrGlyLthr 120  
 DB 344 ACCACGTGCTACAGTACCAACCAAGGCTCTCTATGACGTCACTGCTCATATGCGACT 403  
 QY 121 GlnProAlaTyrProAlaTyrGlyGlnGlnProAlaAlaThrAlaProThrArgProGln 140  
 DB 404 CAGCTGCTTATCCAGCTATGAGGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 463  
 QY 141 AspGlyAanLysProThreGlnThreSerGlnProGlnSerSerThrGlyLysTyrAanGln 160  
 DB 464 GATGAAACAAACCCCACTAGACTAGTCACTCACTCAATACAGCAGGGGTTTCAACAG 523  
 QY 161 ProSerLeuGlyTyrGlyGlnSerAanTyrSerTyrProGlnValProGlySerTyrPro 180  
 DB 524 CCCAGCTAGATATGACAGAGTAACTACAGTATCCCAAGATCTGGAGCTACCC 583  
 QY 181 MetGlnProValThrAlaProProSerTyrProProThraSerTyrSerSerThrGlnPro 200  
 DB 584 ATGCAAGCCAGTACGACCTCATCTACCTCCATCAAGCTATCTCTACACAGCGG 643  
 QY 201 ThrSerTyrAspGlnSerSerTyrSerGlnGlnAanThrTyrGlyGlnProSerSerTyr 220  
 DB 644 ACTAGTATGATCAAGAGCACTTACTCTGCAAGAACCTATGAGGCAACGAGCACTAT 703  
 QY 221 GlyGlnGlnSerSerTyrGlyGlnGlnSerSerTyrGlyGlnGlnProProThraSerTyr 240  
 DB 704 GGCAGCAGAGTATGATGTGTCAACAAAGCACTATGGCAGAGAGCTCCCACTAGTTAC 763



polyA\_signal 2162..2167  
 /gene="EWS"  
 /evidence=experimental  
 polyA\_signal 2350..2355  
 /gene="EWS"  
 /evidence=experimental  
 polyA\_site 2371  
 /gene="EWS"  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 6,25e-77 Length: 2390  
 Score: 3633.00 Matches: 656  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: 9 Gaps: 0  
 US-10-791-017a-2 (1-656) x HSEWS (1-2390)

QY 1 MetAlaSerThrAspTyrSerThrTyrSerGlnAlaAlaGlnGlnGlyTyrSerAla 20  
 Db ATGGCGTCCACGGATTACAGTACCTTATAGCAGCTGCAAGCGCAGCGGCTTACAGTGC 103  
 QY 21 TyrThrAlaGlnProThrGlnGlyTyrAlaGlnThrThrGlnAlaTyrGlnGlnInser 40  
 Db 104 TACACCGCCGACCGCATGAGATATGCAAGACCCAGGCAATATGGGCAACAAAGC 163  
 QY 41 TyrGlyThrTyrGlyGlnProThrAspValSerTyrThrGlnAlaGlnThrThrAlaThr 60  
 Db 164 TATGGAACCTATGAGACAGCCCATGATGTCACTATATCCAGGCTCAGAACCATGCAAC 223  
 QY 61 TyrGlyGlnThrAlaTyrAlaThrThrSerTyrGlyGlnProThrGlyTyrThrThrPro 80  
 Db 224 TATGGGCAAGCCGCTATGACACTTCTTATGACAGCTCCCATGCTTATCTACTCCA 283  
 QY 81 ThrAlaProGlnAlaTyrSerGlnProValGlnGlyTyrGlyThrGlyAlaTyrAspThr 100  
 Db 284 ACTGCCCCCGACGACATACAGCCAGCTGTCCAGGGGTATGGCACTGGTCTTATGATACC 343  
 QY 101 ThrThrAlaThrValThrThrGlnAlaSerTyrAlaAlaGlnSerAlaTyrGlyThr 120  
 Db 344 ACCACGTCTAAGTCAACCAACCCAGGCTCTTATGACAGCTCAGTCTGATATGCACT 403  
 QY 121 GlnProAlaTyrProAlaTyrGlyGlnGlnProAlaAlaThrAlaProThrArgProGln 140  
 Db 404 CAGCCGTCTTATCAGCCTATGGGCAAGCAGCAGCCAGCCCATGCTCAACAAACGCGCAG 463  
 QY 141 AspGlyAsnLysProThrGlnThrSerGlnProGlnSerSerThrGlyGlyTyrAsnGln 160  
 Db 464 GATGGAACCAAGCCCATGAGCTAAGTCAACTCAATCTAGCACAGGGGGTTTCAACCAAG 523  
 QY 161 ProSerLeuGlyTyrGlyGlnSerAsnTyrSerTyrProGlnValProGlySerTyrPro 180  
 Db 524 CCCAGGCTAGATATGACAGAGTAACTACAGTTATCCCAAGGTACTGGGAGACTACCC 583  
 QY 181 MetGlnProValThrAlaProProSerTyrProProThrSerTyrSerSerThrGlnPro 200  
 Db 584 ATGACACGAGTCACTGACCTCATCTTACCTTACCCAGCTATTCCTTACACAGCGG 643  
 QY 201 ThrSerTyrAspGlnSerSerTyrSerGlnGlnAlaThrTyrGlyGlnProSerSerTyr 220  
 Db 644 ACTAGTTATGATAGAGCACTTACTCTCAGCAGAACCTATGGGCAACCGACCACTAT 703  
 QY 221 GlyGlnGlnSerSerTyrGlyGlnGlnInserSerTyrGlyGlnGlnProProThrSerTyr 240  
 Db 704 GGCACACAGAGTACTATGCTCAACAAAGCACTATGGGCAAGCAGCTCCCACTAGTAC 763  
 QY 241 ProProGlnThrThrLysTyrTyrSerGlnAlaProSerGlnTyrSerGlnGlnInserSer 260  
 Db 764 CCAACCCCAAACTGATCTTACAGCCAAAGCTCCAAAGTCAATATAGCCAAAGAACACAGC 823  
 QY 261 TyrGlyGlnGlnInserSerPheArgGlnAspHisProSerSerMetGlyValTyrGlyGln 280

Db 824 TACGGGACAGACAGTTCATCCGACAGACCAACCCAGTACGATGGGTCTTATATGGCAG 883  
 QY 281 GluSerGlyGlyPheSerGlyProGlyGluAsnArgSerMetSerGlyProAspAsnArg 300  
 Db 884 GAGTCTGAGAGATTTTCCCGACAGAGAGAAACCGGAGCATGATGGCCCTGATTAACCCG 943  
 QY 301 GlyArgGlyArgGlyGlyPheAspArgGlyGlyMetSerArgGlyGlyValArgGlyGly 320  
 Db 944 GGCAGGGGAAAGGGGATTTATGTCGAGGCAATGAGAGAGTGGCGGGAGGAGGA 1003  
 QY 321 ArgGlyGlyMetGlySerAlaGlyGluArgGlyGlyPheAsnLysProGlyGlyPromet 340  
 Db 1004 CCGGGTGGAAATGGGACGCGCTGAGAGCGAGGTGGCTTCAATAGCCTGGTGGACCCAG 1063  
 QY 341 AspGlyGlyProAspLeuAspLeuGlyProProValAspProAspGlyLysSerAspAsn 360  
 Db 1064 GATGAAGGACAGACTTGTATCTAGGCTCTCTGTATGATCCAGATGACACTCTGACAA 1123  
 QY 361 SerAlaAlaTyrValGlnGlyLeuAsnAspSerValThrLeuAspAspLeuAlaAspPhe 380  
 Db 1124 AGTGCATTTATGATCAAGAGATTTAATGACAGTGTACTTATGATGATCTGGCAGACTTC 1183  
 QY 381 PheLysGlnCyGlyValValLysMetAsnLysArgThrGlyGlnPrometIleHisIle 400  
 Db 1184 TTTAAGCACTGTGGGGTCTTAAAGTGAACAAGAACTGGGCAACCCATGATCCACATC 1243  
 QY 401 TyrLeuAspLysGlyThrGlyLysProLysGlyAspAlaThrValSerTyrGlyLysPro 420  
 Db 1244 TACCTGACAAAGAAACAGAAAGCCCAAGGCGATGCCAGTGTCTTATGAAAGCA 1303  
 QY 421 PromThrAlaLysAlaAlaValGlnThrPheAspGlyLysAspPheGlnGlySerLysLeu 440  
 Db 1304 CCACTGCGAAGCTGCGCTGGAATGGTTTATGAGGAAGATTTTCAAGGAGCAAACTT 1363  
 QY 441 LysValSerLeuAlaArgLysLysProPrometAsnSerMetArgGlyGlyLeuProPro 460  
 Db 1364 AAAGTCTCCTTGTCTCGAAGAAAGCTCCCAATGAACAGTATGCGGGGTGTGCGCAACC 1423  
 QY 461 ArgGlyGlyArgGlyMetProProProLeuArgGlyGlyProGlyGlyProGlyPro 480  
 Db 1424 CGTGAAGGCAAGGATGCAATGCCACCACTCCGTGAGGTCACAGAGGCCAGAGAGTCT 1483  
 QY 481 GlyGlyProMetGlyArgMetGlyGlyArgGlyGlyAspArgGlyGlyPheProProArg 500  
 Db 1484 GGGGAGCCATGGGTGCAATGGAGAGCCGTGAGAGATAGAGAGGCTTCCCTCCAAAG 1543  
 QY 501 GlyProArgGlySerArgGlyLysAsnProSerGlyGlyGlyAsnValGlnHisArgAlaGly 520  
 Db 1544 GGAACCCGGGGTTCCGAGAGAAACCCCTGAGAGAGAAAGTCCAGACCGAGCTGGA 1603  
 QY 521 AspTyrGlnCysProAsnProGlyCysGlyLysAsnGlnAsnPheAlaThrArgThrGluCys 540  
 Db 1604 GACTGGCACTGTCCCAATCCGGGTGTGGAAACCAAGAACTTGGCCCTGGAGAACAGAGTGC 1663  
 QY 541 AsnGlnCysLysAlaProLysProGlyGlyPheLeuProProProPheProProProGly 560  
 Db 1664 AACCACTGTAAAGGCCCAAGAGCTGAAAGCTTCTCCGCCACCTTTCCGCCCCGGGT 1723  
 QY 561 GlyAspArgGlyArgGlyGlyProGlyGlyMetArgGlyGlyArgGlyGlyLeuMetAsp 580  
 Db 1724 GGTGATCGTGGAGAGGTGGCCCTGGTGCATGCGGGAGGAAGAGTGGCTCATGGAT 1783  
 QY 581 ArgGlyGlyProGlyGlyMetPheArgGlyGlyValArgGlyGlyLysPheArgGly 600  
 Db 1784 CGTGGTGTCTCCGCTGGAAATTTTCAAGGTGGCCGTGGTGGAGACAGAGTGGCTTCCGT 1843  
 QY 601 GlyLysArgGlyMetAspArgGlyGlyPheGlyGlyGlyValArgArgGlyGlyProGlyGly 620  
 Db 1844 GGTGGCGGGGAGTGGACCAAGGTGCTTGTGTGAGGAAGAACAGAGTGGCTTGGGGGG 1903  
 QY 621 ProProGlyProLeuMetGlnGlyMetGlyGlyArgArgGlyGlyValArgGlyGlyProGly 640



[illegible][illegible]



/transa:all:11:"M4ST0Y3T50A0A00G5ATYTA0PT05VACT0A4VG00505VgY  
 G0PFDY20A0T7A4Y0G0A7A1T5G0P20P7A1P0A0NS0P05G0G0A04PTT  
 ATVTTT0A5TA0A05A1G0T0A1YAY0Q0P0A1A1P1R0P0K0K1P0S0P0S0T0G0G0  
 P5L0A0G0S5T1P0V0P0S1P0W0Y0P1A0S1P1T5T50P1S1YD0S1S0Q0N1T0G0S  
 5Y0Q0S1Y0G00S1P0G0P1P5Y0P1P0A5Y0A0P500S5Y0S0S5F0D0H5S0M0S  
 1Y0G0S05F0P0L0E0P0D0P0M0S0P0D0R0G0R0G0P0R0C0M0S0G0G0M0A0G0E0F0P0K  
 P0G0M0D0G0P0L0D0G0P0M0D0S0N0A1Y0Q0L0N5V1L0D1L0A0F0C0G0A0V0K0N0K0T  
 0N0M1K1L0P0R0E0K0P0K0A0T0Y0E0P0P0K0A0W0P0G0P0C0S0K1L0A0R0P0M0  
 G50M0G1L0P0R0E0K0G0M0P1A0G0G0G0G0G0G0G0G0G0G0G0G0G0G0G0G0G0  
 S0G0G0N0H0A0M0G0C0P0R0G0N0N0P0M0R0E0C0C0A0P0A0B0G1L0P0P0P0P0G0D0R0G0G  
 P0M0R0G0R0G0G0G0G0G0K0K0H0E0R0D0P0Y"  
 P0M0R0G0R0G0G0G0G0K0K0H0E0R0D0P0Y"

**ORIGIN**

**Alignment Scores:**

Pred. No.:	1.29e-76	Length:	2189
Score:	3518.50	Matches:	655
Percent Similarity:	99.854	Conservative:	0
Best Local Similarity:	99.854	Mismatches:	0
Query Match:	99.604	Indels:	1
DB:	9	Gaps:	1

US-10-791-017A-2 (1-656) X CR456490 (1-2189)

QY	1	MeLIaSeRThrAPYrSerThrYrSerGlnAlaAlaAGlnGlnIYrSerIa	20
Db	34	ATGGCGTCCAGGATTTCAGTACCTATAGCCAAAGCTGACGGCAGCGGGCTCAAGTGT	93
QY	21	TYrThraIaGlnProthrgInglYrYraIaGlnThrThGlnAlaIaTyrlYgInGlnSer	40
Db	94	TACACGGCCCAAGCCCACTCAAGGATATGCAACACCAAGCCANATGGGCAACAAAGC	15
QY	41	TYrGlyYThrTYrGlyGlnInProthraPvalSerTYrThGlnAlaGlnThrThraIaThr	60
Db	154	TATGGAACTATGAGACAGCCCACTGATGTCACTATACCAAGGCTCAGACCACTCAAC	21
QY	61	TYrGlyGlnInThraIaTYraIaThraPvalSerTYrGlyGlnInProProthGlyYrThrThrPro	80
Db	214	TATGGGAGAGCCCGCTATGCAACTCTCTTATAGACAGCCTCCACGTGGTATCACTCCA	27
QY	81	ThraIaProGlnAlaIaTYrSerGlnProAlaGlnGlyTYrGlyYThrGlyAlaTYraPthr	10
Db	274	ACGCCCCCAAGGCATACAGCCAGCCTGTCCAGGGTATAGCACTGTGCTTATATACC	33
QY	101	ThrThraIaThraValThrThrThrgInAlaSerTYraIaAlaGlnSerAlaTYrGlyYThr	12
Db	334	ACCACTGCTCACTACATCACCAACCCAGGCTCTTATGAGCTCAAGTCGATATAGCACT	39
QY	121	GlnProAlaIaTYrProAlaIaTYrGlyGlnInProAlaAlaIaThraIaProThraPProGln	14
Db	394	CAGCCTGCTTATACAGCCTATAGGAGAGCCAGCCAGCCATGCACTTACAAAGACCGAG	45
QY	141	AspGlyYAsnIaYrProthrgInuThraSerGlnInProGlnSerThrGlyYrYrAaGln	16
Db	454	GATGGAAACAAGCCCACTAGACTAGTCAACTCATCTTACAGCAAGGGGGTATCAACAG	51
QY	161	ProSerIeugIaTYrGlyGlnSerAaNYrSerTYrProGlnValProGlySerTYrPro	18
Db	514	CCCAAGCTTAGAATATGACACGAATACTAACGTTATCCCAAGTACCTGGAGTACCCC	57
QY	181	MetGlnProValThraIaProPProSerTYrProProthSerTYrSerSerThrgInPro	20
Db	574	ATCCAGCAATCACTGACCTCATCTTACCTCCACAGAGCTATTCCTTCAAGACGG	63
QY	201	ThrSerTYrAspIaInSerSerTYrSerGlnGlnIaSerThrTYrGlyGlnProSerSerTYr	22
Db	634	ACTAGTATATATCAAGACAGTATCTTCAAGCAAAACCTATATGGGCAACGAGCAGTAT	69
QY	221	GlyGlnGlnInSerSerTYrGlyGlnGlnInSerSerTYrGlyGlnGlnInProProthSerTYr	24
Db	694	GGACAGACAGATGACATATGCTCAACAAAGCACTATAGGCAAGCCTCCCACTATGTTAC	75
QY	241	ProProGlnInThrGlySerTYrSerGlnAlaProSerGlnTYrSerGlnInSerSerSer	26

Dd		754	CCACCCCAACTGGATTCCTACAGCAGGCTCCAAGTCAATTATGCCAACAAGACAGCAGC	81.3
Oy		261	TyrGlygInGlnSerSerpheArGlnAspPhIspProSerSerMetGlyValTyrGlyGln	280
Dd		814	TACGGGCAGACAGAATTCAATCCGACAGCACACCACCGTAGTGAATGGGTGTTATGGGCAG	87.3
Oy		281	GluSerGlyGlyPheSerGlyProGlyGluAsnArgSerMetSerGlyProAspAsnArg	300
Dd		874	GAGTCGTGAAGGATTTTCCGGACAGAGAGAAACCGGAGCATGATGTGGCTCTGATAACCGG	93.3
Oy		301	GlyArGglYArGglYGLYPheAspArHrgGlyYlmeSerArGlgLyAlrYArHrgGlyYglY	320
Dd		934	GGCAGGGGAAAGGGGGATTATGATCGTGAGGCATAGCAGAGCTGGCGGGGAGAGAGA	99.3
Oy		321	ArgGlyGlyMeGlySerlaGlyGlnArGglYGLYPheAsnlySPProGlyGlyProMet	340
Dd		994	COCGGTGAATGGGC---CGTGAGAGACGAGGTGGCTTCAAATAGCTGTGGAACCCATG	105
Oy		341	AapGlnglyProAspLeuAsplenglYProProValAspProAspGluAspSerAspAsn	360
Dd		1051	GATGAAGACACGACTTGTATGATAGGCCACCTGTATGATCCAGATGAACACTGCAAC	111.1
Oy		361	SerlaletYrValGlnGlyleuAsnAspSerValThlleuAspAspleuIaaApPhe	380
Dd		1111	AGTGCAATTTATGTACAAGATTAATGAACATGTGACTCTGATATATCTGGACACTTC	117.7
Oy		381	PheLysGlnCYSGLYValValLysMetAenlyBARGTHGlYlNPrometIleHISile	400
Dd		1171	TTTAAAGCAGTGTGGGTGTTTAAATGAACAAAGAACTGGGCMAACCTATGATCCACATC	123.3
Oy		401	TyrleuAspLysGluThrGlyLysABroLysGlyAspAlarHrValserTYrGluAspPro	420
Dd		1231	TACCTGGAACAAGAAACAAGAAAGCCCAAGGCCAGATGCCACAGTGTCTTATGAAGACCA	129.9
Oy		421	ProthrAlalyBALalavalGluTrpPheAspGlyLVsAspPheGlnGlySerlySleu	440
Dd		1291	CCCACTGCCAAGGCTGCCGTGGAAATGGTTGATGGGAAAAGATTTTCAAGGGAGCAAACTT	139.9
Oy		441	LysValSerleuAlaArgLysLysProPrometAnSerMetArgGlyGlyLeuProPro	460
Dd		1331	AAAGTCTCCCTTGCTCGAGAGAGGCTTCCAATGAACAGTATGGGGGTGTCTGCCAACC	141.1
Oy		461	ArgGlnglyArGglyMeCProProProleuArGglYGLYProGlyGlyProGlyGlyPro	480
Dd		1411	CGTGAAGGCAGAGGCGATGCCACACCACTCGGTGAGGTCCAAGAGCCCAAGAGGTCTT	147.7
Oy		481	GlyGlyProMetGlyArGmetGlyGlyArGglYGLYAspARGlyGlyPheProProArg	500
Dd		1471	GGGGGACCACATGGGTCTGATGGAGAGCGCTGAGAGAGATAGAGAGGCTTCCCTCCAAGA	153.3
Oy		501	GlyProARgglyserArGglYAsnProserGlyGlyGlyAsnValGlnHisARgalagly	520
Dd		1531	GGACCCCCGGGGTCCCGAGGGGAAACCCCTCTGAGAGAGGAACGTCACAGACCGAGCTGA	159.9
Oy		521	AsPTpGlnCySPCoAsnProGlyCYsglyAsnGlnAsnPheAlatrPARgThrGlnCyS	540
Dd		1591	GACTGSCAGTGTCCCAATCCGGGTGTGTGAACCAAGAACTTGCCCTGGAGAACAGAGTGC	165.5
Oy		541	AsnGlnCyALyBALaProLysProGlnglyPheLeuProProProPheProProGly	560
Dd		1651	AACCAGTGTAAAGCCCCCAAGCCTGAAGGCTTCTCCGCGCACCCCTTCCGCCCCGGGT	171.7
Oy		561	GlyAspARgglyYArGglYGLYProGlyGlyLmeArGglYGLYArGglYGLYleuMetAsp	580
Dd		1711	GGTATCTGTGGCAAGGTGGCCCTGTGTGCATGCGGGGAGGAAGAGGTGGCTCATAGAT	177.7
Oy		581	ArgGlyGlyProGlyGlyMeCPheArGglYGLYArGglYGLYAspARGlyGlyPheArg	600
Dd		1771	CGTGTGTGTCCCGGTGAATGTTCAGAAGTGGCCGTGTGTGAGACAGAGGTGGCTTCCCT	183.3
Oy		601	GlyGlyArGglYmeAspARGlyGlyPheGlyGlyGlyYArGglYGLYProGlyGly	620



QY 141 AspGlyAsnLysProThrGluThrSerGlnProGlnSerSerThrGlyValYAsnGln 160  
 DB 448 GATGGAAACAGCCCACTGAGCTAGCTCACTCACTTACGACAGGGGGTTACACCA 507  
 QY 161 ProSerLeuGlyValGlyGlnSerAsnLysSerThrProGlnValProGlySerThrPro 180  
 DB 508 CCCAGCCTAGGATATGACAGAGTAATCACTACGATTATCCAGAGTACTGGAGAGTACCCC 567  
 QY 181 MetGlnProValThrAlaProProSerThrProProThrSerThrSerThrGlnPro 200  
 DB 568 ATGACAGCACTGACCTGACCTCACTTACCTTACCACTATTCCTCTACACAGCCG 627  
 QY 201 ThrSerThrAspGlnSerSerThrThrSerGlnAsnThrThrGlyGlnProSerSerThr 220  
 DB 628 ACTAGTTATGATCAAGCACTTACTCTCAAGACCACTTATGGCAACCGAGAGCTAT 687  
 QY 221 GlyGlnGlnSerSerThrGlyGlnGlnSerSerThrGlyGlnGlnProProThrSerThr 240  
 DB 688 GGAACAGAGTAGTATGATGATCAACAAAGAGCTATGGCAGAGCTCCACATGTTAC 747  
 QY 241 ProProGlnThrGlySerThrThrSerGlnAlaProSerGlnThrSerGlnGlnSerSer 260  
 DB 748 CCAACCCCAACCTGATCTTACAGCCAACTCAAGTCAATATAGCCAAACAGACAGCAGC 807  
 QY 261 TyrGlyGlnGlnSerSerPheArgGlnAspHisProSerSerMetGlyValThrGlyGln 280  
 DB 808 TACGGGACAGAGTATCTTATCCAGACAGACCCCAAGTACAGAGTGTATATGGGAG 867  
 QY 281 GluSerGlyValLysSerSerGlyProGlyGluAsnArgSerMetSerGlyProAspAsnArg 300  
 DB 868 GAGTCTGAGAGATTTTCCGGACAGAGAGAACCGAGCATGATGGCTCTGATTAACCG 927  
 QY 301 GlyArgGlyValArgGlyValPheAspArgGlyValMetSerArgGlyValArgGlyVal 320  
 DB 928 GGGAGGGGAGAGGGGGATTTGATCGTGAGGATGAGCAGAGTGGCGGGAGAGAGAG 987  
 QY 321 ArgGlyValMetGlySerLysArgGlyValArgGlyValPheAsnLysProGlyValProMet 340  
 DB 988 CGGGGTGAGATGGG---GCTGAGAGAGCGAGTGGCTTCAATTAAGCCTGTGTGAGCCATG 1044  
 QY 341 AspGluGlyProAspLeuAspLeuGlyProProValAspProAspGluAspSerAspAsn 360  
 DB 1045 GATGAAAGACAGATCTTGAATCAAGGCCCACTGATGATCAATGAAGCTTGAACAC 1104  
 QY 361 SerAlaIleThrValGlnGlyLeuAsnAspSerValThrLeuAspAspLeuAlaAspPhe 380  
 DB 1105 AGTGCATTTTATGATCAAGGATTAATGACAGTGTGATGATGATGATGATGATGATG 1164  
 QY 381 PheLysGlnCysGlyValValLysMetAsnLysArgThrGlyGlnProMetIleHisIle 400  
 DB 1165 TTTAAGCAGGTGGGGTGTGTTAAGATGAACAAAGAGATGGGGACCCATGATGATGATG 1224  
 QY 401 TyrLeuAspLysGlnThrGlyLysProLysGlyAspAlaThrValSerThrGluAspPro 420  
 DB 1225 TACCTGCAAGAAACAGAAAGCCCAAGGAGTGCACAGTGTCTTATGAAGCCCA 1284  
 QY 421 ProThrAlaLysAlaAlaValGluTrpPheAspGlyLysAspPheGlnGlySerLysLeu 440  
 DB 1285 CCCACAGCCAAAGGCTCCGTGAAATGTTGATGGAGAAAGATTTTCAAGGAGCAAACTT 1344  
 QY 441 LysValSerLeuAlaArgLysLysProProMetAsnSerMetArgGlyValLeuProPro 460  
 DB 1345 AAAGTCTCCCTGCTCGAAGAGAGCTCCATTAACAGATAGCGGGTGTGTGCAACCC 1404  
 QY 461 ArgGluGlyArgGlyLysProProProLeuArgGlyValProGlyValProGlyValPro 480  
 DB 1405 CCGTGAAGGAGAGAGCATGCCACCACTCCGTGGAGGTCACAGAGGCCCAAGAGTCTCT 1464  
 QY 481 GlyGlyProMetGlyValArgMetGlyValArgGlyValAspArgGlyValPheProProArg 500  
 DB 1465 GGGGAGCCCATGGGTGATGAGAGGCCGTGAGAGAGATAGAGAGAGCTTCCCTCCCAAA 1524

QY 501 GlyProArgGlySerArgGlyAsnProSerGlyValGlyValAsnValGlnHisArgAlaGly 520  
 DB 1525 GGAACCCCGGGGTTCCCGAGGGAAACCTCTTGGAGAGAGAAAGTCCAGCCAGAGCTGGA 1584  
 QY 521 AspTrpGlnCysPheAsnProGlyCysGlyAsnGlnAsnPheAlaTrpArgThrGluCys 540  
 DB 1585 GACTGGCAAGTGTCCCAATCCGGGTTGTGAGAAACAGAACTTGGCTCGAAGAACAGAGTGC 1644  
 QY 541 AsnGlnCysLysAlaProLysProGluGlyPheLeuProProProPheProProGly 560  
 DB 1645 AACCACTGTAAAGCCCAAGCCCTGAAGGCTTCTCCCGCACCTTTCGGCCCGGGGT 1704  
 QY 561 GlyAspArgGlyValArgGlyValProGlyValMetArgGlyValArgGlyValLysMetAsp 580  
 DB 1705 GGTGATCTGGCAGAGAGTGGCTCGTGGATGCGAGTGGAGAGAGAGAGAGAGTGGCTCAAGAT 1764  
 QY 581 ArgGlyValProGlyValLysMetPheArgGlyValArgGlyValAspArgGlyValPheArg 600  
 DB 1765 CGTGTGTGTCCTCGGTGAATGTTCAAGGTGGCCGTGTGAGACAGAGGTGGCTTCCT 1824  
 QY 601 GlyValArgGlyMetAspArgGlyValPheGlyValGlyValArgArgGlyValProGlyVal 620  
 DB 1825 GGTGGCCGGGGCATGAGCCAGAGGTGGCTTGTGTGAGAGAAAGAGAGTGGCCCTCGGGGG 1884  
 QY 621 ProProGlyProLeuMetGluGlnMetGlyValArgArgGlyValArgGlyValProGlyVal 640  
 DB 1885 CCCCCTGAGACCTTGTATGAGAAACAGATGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1944  
 QY 641 LysMetAspLysGlyValLysIleArgGlnGluArgArgAspArgProThr 656  
 DB 1945 AAATGATTAAGCCAGCACCTCTCAGAGCCGAGAGATCGCCCTTAC 1992

RESULT 11  
 AK056309  
 LOCUS  
 DEFINITION  
 Homo sapiens cDNA FLJ31747 f1s, clone NTNR12007377, highly similar  
 to RNA-BINDING PROTEIN EMS.  
 AK056309  
 VERSION  
 AK056309.1 GI:16551673  
 o1go capping; file (full insert sequence).  
 SOURCE  
 Homo sapiens (human)  
 ORGANISM  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
 1  
 Ota, T., Suzuki, Y., Nishikawa, T., Otsuki, T., Sugiyama, T., Irie, R.,  
 Wakamatsu, A., Hayashi, K., Sato, H., Nagai, K., Kimura, K., Makita, H.,  
 Sekine, M., Ohyaishi, M., Nishi, T., Shibahara, T., Tanaka, T.,  
 Ishii, S., Yamamoto, U., Saito, K., Kawai, Y., Isono, Y., Nakamura, Y.,  
 Nagahashi, K., Murakami, K., Yasuda, T., Iwayanagi, T., Wagatsuma, M.,  
 Shiratori, A., Sudo, H., Hosoiri, T., Kaku, Y., Kodaira, H., Kondo, H.,  
 Sugawara, M., Takahashi, M., Kanda, K., Yokoi, T., Furuya, T.,  
 Kikkawa, E., Omura, Y., Abe, K., Kamihara, K., Katsuta, N., Sato, K.,  
 Tanikawa, M., Yamazaki, M., Ninomiya, K., Ishibashi, T., Yamashita, H.,  
 Murakawa, K., Fujimori, K., Tanai, H., Kimura, M., Watanabe, M.,  
 Hirooka, S., Chiba, Y., Ishida, S., Ono, Y., Takiguchi, S., Watanabe, S.,  
 Yosida, M., Hoshino, T., Kusano, Y., Kanehori, K., Takahashi, F.,  
 Hara, H., Tanabe, T., Nomura, Y., Togiya, S., Komai, F., Hara, R.,  
 Takeuchi, K., Arita, M., Imose, N., Mutsaers, K., Yuuki, H., Oshima, A.,  
 Saeki, N., Aocsa, S., Yoshikawa, Y., Matsumura, H., Ichihara, T.,  
 Shiohara, N., Sano, S., Moriya, S., Momiyama, H., Satoh, N., Takami, S.,  
 Terashima, Y., Suzuki, O., Nakagawa, S., Senoh, A., Mizoguchi, H.,  
 Goto, Y., Shimizu, F., Wakebe, H., Hishigaki, H., Watanabe, T.,  
 Sugiyama, A., Takemoto, M., Kawakami, B., Yamazaki, M., Watanabe, K.,  
 Kumagai, A., Itakura, S., Fukuzumi, Y., Fujimori, Y., Komiyama, M.,  
 Tashiro, H., Tanigami, A., Fujisawa, T., Ono, T., Yamada, K., Fujii, Y.,  
 Ozaki, K., Hizo, M., Ohmori, Y., Kawabata, A., Hikiji, T., Kobatake, N.,  
 Inagaki, H., Ikema, Y., Okamoto, S., Ohtani, R., Kawakami, T.,  
 Noguchi, S., Itoh, T., Shigeta, K., Senba, T., Matsumura, K.,  
 Nakajima, Y., Mizuno, T., Morinaga, M., Saeki, M., Togaishi, T.,  
 Oyama, M., Hata, H., Watanabe, M., Komatsu, T., Mitsuishi-Sugano, J.,  
 Satoh, T., Shirai, Y., Takahashi, Y., Nakagawa, K., Okumura, K.,  
 Nagase, T., Nomura, N., Kikuchi, H., Masuno, Y., Yamashita, R.,







Db 1099 GATGAAGACTCTGCAACAGTGCATTTATGTAACAAGATTAATGACAGTGTGACTCTTA 1158  
 Qy 375 AAPPAPLEUAlAaPhePheLySgInCySgLyValValLyMeLanLyBaTgThrGly 394  
 Db 1159 GATGACTGCGAGACTCTTTAAAGCAGTGTGGGTGTTAAAGTGAACAAGAACCTGGG 1218  
 Qy 395 GlnProMetIleHisIleTyrLeuAaPlySgLuThGlyLyBaProLySgIyaPaLaTtr 414  
 Db 1219 CAACCATATATCCATCTACTGAGAACGAAACGAAAGCCCAAGGCGATGCCACA 1278  
 Qy 415 ValSerTyrGlyAaPProProThraIaLyBaIaIaValGluTrrPheAaPglyLyBaP 434  
 Db 1279 GTGTCTTATGAAGACCCACCACTGCCAAGGCTGCCGTGGAAAGTTTATGGGAAGAT 1338  
 Qy 435 PheGInGlySerLyLeuLyValSerLeuAlaTgLySgLyBaProProMeLanSerMet 454  
 Db 1339 TTTCAAGGAGAGCAAACTTAAGCTCTCCCTGTGCTCGAAGAACCTCAATGAACAGTATG 1398  
 Qy 455 ATGSGLyLyLeuProProAaTgLyAaTgLySgLyMeProProProLeuAaTgLySgLy 474  
 Db 1399 CGGGTGTGTCTGCACCCGCTGAGGAGGAGGATGCAACCACTCCGTGGAGGTCCA 1458  
 Qy 475 GLyGlyProGlyLyProGlyLyProMetGlyAaTgMetGlyLyAaTgLySgLyAaPax 494  
 Db 1459 GGAAGGCCAGAGAGTCTCTGGGGAGCCCATGGGTGCGATGGAGGCCGTGGAGAGATAGA 1518  
 Qy 495 GLyGlyPheProProAaTgLyProAaTgLySerAaTgLyAaPProSerGlyLySgLyAaP 514  
 Db 1519 GGAAGGCTTCCCTCCAAAGAGACCCCGGGGTTCCTCGAGGGAACCCCTCGAGAGAGGAAC 1578  
 Qy 515 ValGInHisAaPAlaGlyAaPTrpGInCyBaProAaPProGlyCyBaGlyAaPInaPhe 534  
 Db 1579 GTCCAGACCGAGCTGAGAGCTGGCAGTGTCCCAATCCGGGTGTGAAACAGAACTTC 1638  
 Qy 535 AlaTyrAaTgThrGInCyBaPInaPheAaPProGlyLyAaPProGlyLyPheLeuProPro 554  
 Db 1639 GCCTGAGAAACAGAGTGAACCAAGTGAAGGCCCAAGGCTGAAGGCTTCCTCCGCCCA 1698  
 Qy 555 ProPheProProProGlyLyAaPProGlyLyAaTgLySgLyMeLanLyBaTgThrGly 574  
 Db 1699 CCGTTTCCCGCCCCCGGGGTGTATCGTGGCAGAGTGGCCCTGTGGCATGCGGGAGAGA 1758  
 Qy 575 ATGSGLyLyLeuMetAaPProGlyLyAaTgLySgLyMeLanLyBaTgThrGly 594  
 Db 1759 AAGAGGTGCTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1818  
 Qy 595 AaPProGlyLyPheAaTgLySgLyAaTgLyMeLanLyBaTgThrGlySgLyAaTg 614  
 Db 1819 GACAGAGTGTGCTTCCGTGTGTGCTGCGGCGCATGGAACGAGTGTGTTGTGGAGAGA 1878  
 Qy 615 ATGSGLyLyProGlyLyProProGlyLyProLeuMetGInGInMetGlyLyAaTgAaTg 634  
 Db 1879 CAGAGGTGCTCTGGGGGGCCCCCTTGACCTTTGATGAAACAGTGGAGAGAAAGAGA 1938  
 Qy 635 GLyAaTgLySgLyProGlyLyMeLanLyBaTgLySgLyHisAaTgGInGInAaTgAaPax 654  
 Db 1939 GAGCGTGGAGAGCTGGAAAAATGATAAAGGAGACCGCTCAGAGGCGCAGAGATCGG 1998  
 Qy 655 ProTyr 656  
 Db 1999 CCTTAC 2004  
 RESULT 12  
 CQ730352 2326 bp DNA linear PAT 03-FEB-2004  
 LOCUS CQ730352 16286 from Patent WO02068579.  
 DEFINITION CQ730352  
 ACCESSION CQ730352  
 VERSION CQ730352.1 GI:42303963  
 KEYWORDS  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1  
 AUTHORS Venter, C.J., Adams, M.C., Li, P.W. and Myers, E.W.  
 TITLE Kites, such as nucleic acid arrays, comprising a majority of  
 humenexons or transcripts, for detecting expression and other uses  
 thereof  
 JOURNAL Patent: WO 02068579-A 16286 06-SEP-2002;  
 FEATUES PE Corporation (NY) (US)  
 source location/Qualifiers  
 1. 2326  
 /organism="Homo sapiens"  
 /mol\_type="unassigned DNA"  
 /db\_xref="taxon:9606"  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 3,22e-76 Length: 2326  
 Score: 3602.00 Matches: 652  
 Percent Similarity: 99.85 Conservative: 0  
 Best Local Similarity: 99.85 Mismatches: 0  
 Query Match: 99.15 Indels: 1  
 DB: Gaps: 0  
 US-10-791-017A-2 (1-656) x CQ730352 (1-2326)  
 Qy 5 AAPPYrSerThrTyrSerGlnAlaAlaGlnGlnGlyTyrSerAlaTyrThrAlaGln 24  
 Db 8 GATTACAGTACTTATAGCCAGCTGACGCGCAGCAGGAGGCTTACAGCTTACACCGGCCAG 67  
 Qy 25 ProThGInGlyTyrAlaGlnThrThrGlnAlaTyrGlyGlnGlnSerTyrGlyTyrTyr 44  
 Db 68 CCACTCAAGATATGACAGACACCAAGGATATGGGCAACAAAGCTATGGAACCTAT 127  
 Qy 45 GLyGInProThraPyaLySerTyrThrGlnAlaGlnThrThraThrTyrGlyGInThr 64  
 Db 128 GGAAGCCCACTGATGTCAGTATACCCAGGCTCAGACCACTGCAACCTATGGGAGACC 187  
 Qy 65 AlaTyrAlaThrSerTyrGlyGlnProProThrGlyTyrThrProThraIaProGln 84  
 Db 188 GCTATGCACTTCTTATGAGACGCTCCACCTGTTATATCTCAACTGCCCCCAAG 247  
 Qy 85 AlaTyrSerGlnProValGlnGlyTyrGlyTyrGlyTyrAlaTyrThrThraIaThr 104  
 Db 248 GATACAGCAAGCTCTTCCAGGGATGAGCACTGGCTTATGATACACCACTGCTACA 307  
 Qy 105 ValThrThrGlnAlaSerTyrAlaAlaGlnSerAlaTyrGlyThrGlnProAlaTyr 124  
 Db 308 GTACACCAACCAAGGCTCTATGACGCTCAGTGTGATATGGCACTGAGCTGCTTAT 367  
 Qy 125 ProAlaTyrGlyGlnGlnProAlaAlaThrAlaProThraPaxProGlnAaPlyAaPly 144  
 Db 368 CAGGCTTATGGCAGAGCAGCAGCACTGACCTTCAACAGACCGAGAGTGGAAACAAG 427  
 Qy 145 ProThrGlnThrSerGlnProGlnSerSerThrGlyTyrTyrAaPProSerLeuGly 164  
 Db 428 CCACTGAGACTATGTCATCTTATCTAGACAGGGGGTTACACCAAGCCCACTGAGA 487  
 Qy 165 TyrGlyGlnSerAaPlySerTyrProGlnValProGlySerTyrProMetGlnProVal 184  
 Db 488 TATGACAGAGTAACTACAGTATATCCCAAGTAACTGGAGGCTACCCCATGAGCCAGTC 547  
 Qy 185 ThrAlaProProSerTyrProProThraSerTyrSerSerThrGlnProThraSerTyrAaP 204  
 Db 548 ACTGACCTTCACTTCACTTCACTTCACTTCACTTCACTTCACTTCACTTCACTTCACTT 607  
 Qy 205 GlnSerSerTyrSerGlnGlnAaPThrTyrGlyGlnProSerSerTyrGlyGlnGlnSer 224  
 Db 608 CAGAGCAATTACTTCAAGCAAGCACTTATGGCAACCAAGCAAGCTATGAGACAGCAAGT 667  
 Qy 225 SerTyrGlyGlnGlnSerSerTyrGlyGlnGlnProProThraSerTyrProProGlnThr 244  
 Db 668 ACCTATGTGCAACAAAGCAGCTATGGGAGAGGCTCCCACTATGTTACCAACCCCAACT 727  
 Qy 245 GlySerTyrSerGlnAlaProSerGlnTyrSerGlnGlnSerSerSerTyrGlyGlnGln 264

Db 728 GGAATCTAAGCCAGACCTCAATCAATTAAGCAACAGAGCAGCTACCGGACGCG 787  
 Qy 265 SerSerPheArgGlnAphIspSerSerMetGlyValTyrGlyGlnGlnSerGlyGly 284  
 Db 788 AGTTCTTCCGACAGAGCAACCCCACTGATGCTGGTGTATGAGGAGAGCTGAGAGA 847  
 Qy 285 PheSerGlyProGlyGlnAphAphSerMetSerGlyProAphAphArgGlyTyrGly 304  
 Db 848 TTTTCCGACAGAGAGAAACCGAGACATGAGTGGCTGATTAACGGGACAGGAGAGA 907  
 Qy 305 GlyGlyPheAphArgGlyGlyMetSerArgGlyGlyTyrGlyGlyGlyGlyMet 324  
 Db 908 GGGGAGTTTGTATGCTGAGGAGCAATGACAGAGTGGGCGGAGAGAGAGCGGTGGAAG 967  
 Qy 325 GlySerAlaGlyGlyGlnArgGlyGlyPheAphAphProGlyGlyProMetAphGlyPro 344  
 Db 968 GCGAGCGCTGAGAGAGGAGGTGGCTTCAATTAAGCTGTGAGCCATGATGAGAGCA 1027  
 Qy 345 AApLeuAphLeuGlyProProValAphProAphGlyAphSerAphAphSerAlaIleTyr 364  
 Db 1028 GATCTTGAATCTAGGCCCCACCTGTAGATCCAGATGAAGACTGTGACAAAGTGCATTTAT 1087  
 Qy 365 ValGlnGlyLeuAphAphSerValThrLeuAphAphLeuAlaAphPhePheGlyGlnCys 384  
 Db 1088 GTTCAAGATTAATGACAGTGTGACTTATGATGATCTGGCAGACTTCTTTAAGCAGTGT 1147  
 Qy 385 GlyValValIleMetAphAphAphAphAphAphAphAphAphAphAphAphAphAph 404  
 Db 1148 GGGGTGTATTAAGATGAAGAAAGAAAGTGGGCAACCATGATCCATCACTGAGGCAAG 1207  
 Qy 405 GlnThrGlyAphAphProGlyAphAphAphAphAphAphAphAphAphAphAphAphAph 424  
 Db 1208 GAAGACGAGAAAGCCCAAGGCGATGCGACAGTGTCTTATGAGACCAACCACTGCGCAAG 1267  
 Qy 425 AlaAlaValGlnTrpPheAphAphAphAphAphAphAphAphAphAphAphAphAph 444  
 Db 1268 GCTGCGGTGGAAGTGTGATGAGGAAAGATTTTCAAGGAGGCAAACTTAAAGTCTCCTT 1327  
 Qy 445 AlaArgGlyAphAphAphAphAphAphAphAphAphAphAphAphAphAphAphAphAph 464  
 Db 1328 GCTCGGAAAGAGCTCAATGAACATATGCGGGGTGTCTGTCACACCCGCTGAGGCAAG 1387  
 Qy 465 GlyMetProProProLeuAphAphAphAphAphAphAphAphAphAphAphAphAph 484  
 Db 1388 GGGATCCACCAACCACTCCGTGAGGAGTCCAGAGGCGCCAGAGAGTCTTGGGAGACCAAG 1447  
 Qy 485 GlyAphMetGlyGlyTyrArgGlyGlyAphAphAphAphAphAphAphAphAphAphAph 504  
 Db 1448 GGTGCGATGAGAGCGCTGAGAGAGATGAGAGAGGCTTCTCCCAAGAGAGACCCGGGGT 1507  
 Qy 505 SerArgGlyAphAphAphAphAphAphAphAphAphAphAphAphAphAphAphAphAph 524  
 Db 1508 TCCCGAGGAGAAACCCCTCTGAGAGAGAGAAAGTCCACACCAAGACTGAGAGTGGCAGTGT 1567  
 Qy 525 ProAphAphAphAphAphAphAphAphAphAphAphAphAphAphAphAphAphAphAph 544  
 Db 1568 CCCAAATCCGGGTGTGAGAAACCAAGAACTTCTGAGAGAGAGAGAGAGAGAGAGAGAGAG 1627  
 Qy 545 AlaAphAphAphAphAphAphAphAphAphAphAphAphAphAphAphAphAphAphAph 564  
 Db 1628 GCCCAAAAGCTTAAGGCTTCTCCCGCCACCTTTCGCGCCCGGGGTGTGATCTGTGGG 1687  
 Qy 565 ArgGlyGlyProGlyGlyMetAphAphAphAphAphAphAphAphAphAphAphAphAph 584  
 Db 1688 AGAGGTGGCCCTGTGAG 1747  
 Qy 585 GlyGlyMetPheAphAphAphAphAphAphAphAphAphAphAphAphAphAphAphAphAph 604  
 Db 1748 GGTGAGATGTTGAGAGGTGGCGGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1807  
 Qy 604 YMetAphAphAphAphAphAphAphAphAphAphAphAphAphAphAphAphAphAphAph 624

Db 1808 CATGACCGAGGTGCTTTGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1867  
 Qy 624 OleMetGlnGlnMetGlyTyrArgArgGlyGlyTyrArgGlyGlyProGlyTyrMetAphAph 644  
 Db 1868 TTTGATGAGAAAG 1927  
 Qy 644 GGIYGLHISARGINGIUAARGARGAPSPROTYR 656  
 Db 1928 AGCGAGACACCTTCAG 1964  
 RESULT 13  
 BC068226 2189 bp mRNA linear ROD 23-AUG-2004  
 LOCUS Mus musculus Bwing sarcoma breakpoint region 1, mRNA (CDNA clone  
 DEFINITION MG:61279 IMAGE:5707017), complete cds.  
 ACCESSION BC068226  
 VERSION BC068226.1 GI:46399228  
 KEYWORDS MGC.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 2189)  
 Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G.,  
 Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D.,  
 Altschul, S.F., Zeeberg, B., Buetow, K.H., Scheffer, C.F., Bhat, N.K.,  
 Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F.,  
 Diatchenko, L., Marusheva, K., Farmer, A.A., Rubin, G.M., Hong, L.,  
 Stadelton, M., Soares, M.B., Bonaldi, M.F., Casavant, T.L.,  
 Scheetz, T.E., Brownstein, M.J., Uebli, T.B., Toshiyuki, S.,  
 Carninci, P., Prange, C., Rana, S.S., Loquellano, N.A., Peters, G.J.,  
 Abramson, R.D., Mullan, S.J., Bosak, S.A., McMan, P.J.,  
 McKernan, K.J., Malek, J.A., Gamarale, P.H., Richards, S.,  
 Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W.,  
 Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,  
 Fahey, J., Helton, E., Kettelman, M., Madan, A.C., Rodriguez, S.,  
 Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y.,  
 Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D.,  
 Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmitt, J., Myers, R.M.,  
 Butterfield, I.S., Krzywicki, M.I., Skalska, U., Smal, D.B.,  
 Schnerch, A., Schein, J.E., Jones, S.J., and Marra, M.A.  
 Generation and initial analysis of more than 15,000 full-length  
 human and mouse cDNA sequences  
 Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)  
 12477932  
 2 (bases 1 to 2189)  
 Director MGC Project.  
 Direct Submission  
 Submitted (30-MAR-2004) National Institutes of Health, Mammalian  
 Gene Collection (MGC), Cancer Genomics Office, National Cancer  
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
 USA  
 NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
 CONTACT: MGC help desk  
 Email: [cgabbs-remail.nih.gov](mailto:cgabbs-remail.nih.gov)  
 Tissue Procurement: Dr. Jim Ikin, University of Iowa  
 CDNA Library Preparation: M. Bento Soares, University of Iowa  
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNU)  
 DNA Sequencing by: University of Iowa, Dr. M. Bento Soares and Dr.  
 Thomas L. Casavant.  
 Web site: <http://genome.uiowa.edu>  
 Contact: [bento-soares@uiowa.edu](mailto:bento-soares@uiowa.edu); [tom-casavant@uiowa.edu](mailto:tom-casavant@uiowa.edu)  
 Bonaldo, M.F., Akabogu, I., Bair, T., Bair, J., Crouch, K., Davis, A.,  
 Fisher, K., Keppel, C., Kucaba, T., Lebeck, M., Melo, A., Schaefer, K.,  
 Scheetz, T., Smith, C., Smith, E., Tack, D., Trout, K., Walters, J.,  
 Casavant, T., Soares, M.B.  
 Clone distribution: MGC clone distribution information can be found  
 through the I.M.A.G.E. Consortium/LNU at: <http://image.lnl.gov>  
 Series: Plate: Row: Column: 0  
 This clone was selected for full length sequencing because it  
 passed the following selection criteria: matched mRNA gi: 6679714.  
 Location/Qualifiers  
 FEATURES



QY 501 GYPRORAGLYSeraRGLYanProSerGlyYGLYVAnValGlnHsaArgLaGly 520  
 DB 1535 GGGCCCCGAGGCTCCAGAGAAAACCCCTGAGAGAGAAATGCCAGACCGAGCTGGA 1594  
 QY 521 AAPTPTGInCyPrOanProGlyCySGLYAnGlnAnPheAlaTParGThrGluCyS 540  
 DB 1595 GACTGGCACTGTCCCAATCCGGGCTGTGAAACAGAACTTCGCTTGAGAAACAGATTC 1654  
 QY 541 AANGInCyELyALaProLyProGInGlyPheLeuProProPheProProGly 560  
 DB 1655 AACCAAGTAAAGCCCTTAAGCCCGAGGCTTCTCCGCCACCTTCCTCCGGGT 1714  
 QY 561 GYAAPARAGLYARGLYGLYProGlyYGLYMeARGLYGLYARGLYGLYLeuMeAR 580  
 DB 1715 GGTGATCGGAGAGAGAGTGCCTGGTGCATGCGAGAGAGAGAGAGAGTCAAGAC 1774  
 QY 581 ATGGLYGLYProGlyYGLYMeARGLYGLYARGLYGLYVAPARGLYGLYPheAR 600  
 DB 1775 CGTGGTGTCTCGAGAGAAATGTTCAAGGTGCGAGAGTGGAGACAGAGAGCTTCCGA 1834  
 QY 601 GYGLYARGLYMeARGLYGLYPheGlyYGLYARGLYGLYARGLYGLYProGly 620  
 DB 1835 GGTGGCCGTGAATGAGACGAGGTGCTTGTGTGAGAGAAAGACAGAGTGTCTCGGGGG 1894  
 QY 621 PROProGlyProLeuMeTGLuGlnMeTGLYGLYARGLYGLYARGLYGLYProGly 640  
 DB 1895 CCTCTGAGACCTTAATGAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1954  
 QY 641 LYMeARGLYGLYGLYHsaRGLYGLYARGLYGLYARGLYGLYARGLYGLYProGly 656  
 DB 1955 AAAATGATTAAGCGAGACCTCTCAGAAACGAGAGAGAGAGAGAGAGAGAGAGAGAG 2002

RESULT 14  
 AX305537  
 LOCUS Sequence 288 from Patent WO0188188.  
 DEFINITION AX305537  
 ACCESSION AX305537  
 VERSION AX305537.1 GI:117645025  
 KEYWORDS  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus  
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.

REFERENCE 1  
 Ieshikawa, K., Arai, S., Takahashi, Y., Negata, T. and Ishii, Y.  
 Method for examining ischemic conditions  
 Patent: WO 0188188-A 288 22-NOV-2001;  
 School Juridical Person Nihon University (JP)

FEATURES  
 source location/Qualifiers

1..2188  
 /organism="Mus musculus"  
 /mol\_type="unassembled DNA"  
 /db\_xref="taxon:10090"

## ORIGIN

Alignment Scores:  
 Pred. No.: 1.58e-75 Length: 2188  
 Score: 3571.50 Matches: 643  
 Percent Similarity: 99.24% Conservative: 8  
 Best Local Similarity: 98.02% Mismatches: 4  
 Query Match: 98.31% Indels: 1  
 DB: 6 Gaps: 1

US-10-791-017a-2 (1-656) x AX305537 (1-2188)

QY 1 MetAlSeThrApYrSeThrTYrSerGlnAlaAlaGlnGlnGlnTYrSerAla 20  
 DB 57 ATGGCTTCACGGATTACGATACCTATATGACAGCTGACCCAGCAGGGCTCAAGTCT 116  
 QY 21 TYrThAlaGlnProThrGlnGlyTYrAlaGlnThrThrGlnAlaTYrGlnGlnInser 40  
 DB 117 TTAACGGCCAGCCAACTCAAGATATGACAGACCAAGGCAATATGGCAACAAAG 176

QY 41 TYrGlyThrTYrGlnProThrApValSerTYrThrGlnAlaGlnThrThrAlaThr 60  
 DB 177 TATGAACTTATGAGACCTACTGATCTACCTTACTAGGCTCAGACCACTGCCACC 236  
 QY 61 TYrGlnThrAlaTYrAlaThrSerTYrGlnProProThrGlyTYrThrPro 80  
 DB 227 TACGGGAGACTGATATGCAACTCTTACGGACAGCTCCCACTGGTTATAGTACTCCA 296  
 QY 81 ThrAlaProGlnAlaTYrSerGlnProValGlnGlyTYrGlyThrGlyValaTYrAspThr 100  
 DB 297 ACTGCCCCAGGCTTACAGCAGGCTGTGAGAGGATATGCACTGGGGCTTATACAGAC 356  
 QY 101 ThrThrAlaThrValThrThrThrGlnAlaSerTYrAlaAlaGlnSerAlaTYrGlyThr 120  
 DB 357 ACCACTGCTACGTATCAACAGAGAGGCTCTTACGAGCTCACTACAGCATATGGCACC 416  
 QY 121 GlnProAlaTYrProAlaTYrGlyGlnInProAlaAlaThrAlaProThrArgProGln 140  
 DB 417 CAGCTGCTACCCACCTATGGCCAGAGCCAAAGCCAGCCAGCTTACAGACCAAG 476  
 QY 141 AspGlyAnuLyProThrGlnThrSerGlnProGlnInserSerThrGlyTYrAnGln 160  
 DB 477 GATGTAAACAGCTCTGAGACTGACCACTCAATCTAGCAGAGGGGTTATACCA 536  
 QY 161 ProSerLeuGlyTYrGlnInserAnTYrSerTYrProGlnValProGlySerTYrPro 180  
 DB 537 CCAGCTTATGATATGAGACAGATTAAGTACTTACCTTCCAGGTACTCTGGAGCTACCA 596  
 QY 181 MetGlnProValThrAlaProProSerTYrProProThrInserTYrSerSerThrGlnPro 200  
 DB 597 ATGAGCCAGTACCGCACTCCTCACTTATCTCTTACAGGTACTCTCTTACAGCCG 656  
 QY 201 ThrSerTYrAspGlnInserSerTYrSerGlnGlnAnThrTYrGlnProSerSerTYr 220  
 DB 657 ACTATTACGATCAAGAGATTAAGTACTCTGAGCAACCTTATGGGAGCGAGAGCTAT 716  
 QY 221 GYGLYGLYInserSerTYrGlnInserSerTYrGlnGlnInProProThrSerTYr 240  
 DB 717 GAGCAACAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 776  
 QY 241 PROProGlnThrGlySerTYrSerGlnAlaProSerGlnTYrSerGlnInserSerSer 260  
 DB 777 CGGCTCAGACTGATCTTACAGCCAGGCTCCAAATATATAGCAACAGAGCAGCAGC 836  
 QY 261 TYrGlyGlnInserSerPheArgGlnInserPheProSerSerMetGlyValTYrGlyGln 280  
 DB 837 TACGGGAGAGAGATTAAGTACTCTGAGCAACCTTATGGGAGCTTATGGGAG 896  
 QY 281 GlnSerGlyGlyPheSerGlyProGlyGlnAnArgSerMetSerGlyProAspAnArg 300  
 DB 897 GATCTGAGAGATTTTCCGACCAAGAGAAACCGAGCTTATAGTCCCTGATTAACCG 956  
 QY 301 GYAARGLYARGLYGLYPheARGLYGLYMeARGLYGLYARGLYGLYARGLYGLY 320  
 DB 957 GGCAGGGAGAGAGGGGATTTGATCGTGAAGCATGAGCAGAGTGGGGAGAGAGGA 1016  
 QY 321 ArgGlyGlyMeGlySerAlaGlyLuarGlyGlyPheAnuLyProGlyProMet 340  
 DB 1017 CGCGGTGAGCTGGC---GCTGAGAGAGAGAGGCTTCAATAGCTGTGGTGAACCAAG 1073  
 QY 341 AspGlnGlyProAspLeuAspLeuGlyProProValAspProAspGlyLuarAspAn 360  
 DB 1074 GATGAGAGCCAGATCTTATAGCTTACGCTTCTTATAGTCCGAGTACAGACTTGAAC 1133  
 QY 361 SerAlaIleTYrValGlnGlyLeuAnAspSerValThrLeuAspAspLeuAlaAspPhe 380  
 DB 1134 AGTGAATTTATGTGAAGAGTTAATATGACATATGAGCTCTGAGATCTGGAGACTTC 1193  
 QY 381 PheLYeGlnCyGLYValaValaLYeMeCAnuLYeArgThrGlyGlnProMetIleHisIle 400  
 DB 1194 TTTAAGCACTGTGGGTGTCTCAAGATGAACAAAGAGAACTGCAACCACTGATCATATTC 1253



```

Db 477 GATGTTAAAGAGCTGCTGAGCTAGCAACCTCAATCTAGCAAGGGGGTTATTAACCA 536
Qy 161 ProSerLeuGlyTyrGlyGlnSerAsnTyrSerTyrProGlnValProGlySerTyrPro 180
Db 537 CCCAGGCTAGGATATGAGACAGAGTAATCTACAGCTATCCCAAGTACCTGGAGCTTACCA 596
Qy 181 MetGlnProValThrAlaProProSerTyrProProThrSerTyrSerSerThrGlnPro 200
Db 597 ATGAGCAAGTACACCGACCTTCACTTATCTTACAGCTACTCTCTTCAAGCCG 656
Qy 201 ThrSerTyrAspGlnSerSerTyrSerGlnGlnAsnThrTyrGlyGlnProSerSerTyr 220
Db 657 ACTAGTTAGATCAGAGCACTTACTCTCAGCAGAACCTATGGGAGCCGACAGCTAT 716
Qy 221 GlyGlnGlnSerSerTyrGlyGlnGlnSerSerTyrGlyGlnGlnProProThrSerTyr 240
Db 717 GGAACAAGAGTACTATGTCACAAACAGCTATGGGAGCAGCTCTCTACTAGTTAC 776
Qy 241 ProProGlnThrGlySerTyrSerGlnAlaProSerGlnTyrSerGlnGlnSerSer 260
Db 777 CCGCTTCAAGCTGATCTTACAGCCAGGCTTCAAGTCAATATAGCCAAACAGACAGCAGC 836
Qy 261 TyrGlyGlnGlnSerSerPheArgGlnAspHisProSerSerMetGlyValTyrGlyGln 280
Db 837 TAGGGGAGAGAGTTTCCGACAGGACCAACCCAGTAGCATTGGGTGTTATGGGAG 896
Qy 281 GluSerGlyGlyPheSerGlyProGlyGlyLysAsnArgSerMetSerGlyProAspAsnArg 300
Db 897 GAGCTGAGAGATTTTCCGACAGGACCAACCCAGTAGCATTGGGTGTTATGGGAG 956
Qy 301 GlyArgGlyValArgGlyPheAspArgGlyGlyMetSerArgGlyGlyArgGlyGly 320
Db 957 GGCAGGGGAGAGGGGGATTTGATCTGAGAGCATGAGCAGAGTGGGCGGGAGAGAG 1016
Qy 321 ArgGlyGlyMetGlySerAlaGlyLysArgGlyGlyPheAsnLysProGlyGlyProMet 340
Db 1017 CGCGGTGAGCTGGGCG---GCTGGAGAGCGAGGTGGCTTCAATTAAGCTGTGAGCAACCATG 1073
Qy 341 AspGlyGlyProAspLeuAspLeuGlyProProValAspProAspGlyLysAspAsn 360
Db 1074 GATGAAGAGACCAATCTTGAATCAAGGCTTCTTAATGATCCCATGAAGACTCTGACAAAC 1133
Qy 361 SerAlaIleTyrValGlnGlyLysAsnAspSerValThrLeuAspLeuAlaAspPhe 380
Db 1134 AGTGCATTTATGTGACAGATTAATGACATGTGACTGTGATGATCTGGCACTTC 1193
Qy 381 PheLysGlnCysGlyValValLysMetAsnLysArgThrGlyGlnProMetIleHisIle 400
Db 1194 TTTAAGCAGTGTGGGCTTCAAGATGAACAAGAGACTGACAAACCATGATCCATATC 1253
Qy 401 TyrLeuAspLysGlyThrGlyLysProLysGlyAspAlaThrValSerTyrGlyLysPro 420
Db 1254 TACCTGATTAAGAGACAGAGAAAGCTTAAAGGGGAGCCCAAGTGTCTTATGAAGATCCA 1313
Qy 421 ProThrAlaLysValAlaValGluTyrPheAspGlyLysAspPheGlnGlySerLysLeu 440
Db 1314 CCAACTCAAAAGCTGCGCTGGAATGTTGATGGGAAAGATTTTCAAGGAAGCAAACTT 1373
Qy 441 LysValSerLeuAlaArgLysLysProProMetAsnSerMetArgGlyLysLeuProPro 460
Db 1374 AAAGTGTCTTGGCCGAAAGAGCTTCAATGAAACAGATGCGGGGAGCATGCCACT 1433
Qy 461 ArgGlyGlyArgGlyMetProProProLeuArgGlyGlyProGlyGlyProGlyGlyPro 480
Db 1434 CGTGAAGGAGAGGGGTATGCAACCACTTGTGAGGTCTGTGTGCTCCAGAGAGGCTT 1493
Qy 481 GlyGlyProMetGlyArgMetGlyArgGlyValAspArgGlyGlyPheProProArg 500
Db 1494 GAGAGACCTCAATGGTCCCAATGGAGGCGCTGAGAGAGACAGAGGGGCTTCCCTCCAAAG 1553
Qy 501 GlyProArgGlySerArgGlyAsnProSerGlyGlyGlyLysValGlnIleArgAlaGly 520

```

```

Db 1554 GGGCCCGAGGCTCCAGAGAAACCTCTGAGGAGGAAATGTCCAGACCCAGCTGGA 1613
Qy 521 AspTyrGlnCysProAsnProGlyCysGlyValGlnGlnAsnPheAlaThrArgThrGlyCys 540
Db 1614 GACTGACAGTGTCCCAATCCGGGCTGTGAAACCAAGAACTTGTGGAGAAACAGATGC 1673
Qy 541 AsnGlnCysLysAlaProLysProGlyGlyPheLeuProProProPheProProGly 560
Db 1674 AACCAAGTAAAGGCCCTTAAGCCGAGGGCTTCTCCGCGCAACCTTTCACCTCGGGT 1733
Qy 561 GlyAspArgGlyValArgGlyProGlyGlyMetArgGlyGlyValArgGlyGlyLysMetAsp 580
Db 1734 GGTGATCGTGAAGAGGTGGCCCTGTGTGATGCAAGAGAGAGAGAGAGACTCATGAGC 1793
Qy 581 ArgGlyGlyProGlyGlyMetPheArgGlyValArgGlyGlyValAspArgGlyGlyPheArg 600
Db 1794 CGTGTGTCTGTGAGAAATGTTCAAGAGTGCAGAGGTGGAGACAGAGAGAGGCTTCCGA 1853
Qy 601 GlyGlyArgGlyMetAspArgGlyGlyPheGlyGlyGlyArgArgGlyGlyProGlyGly 620
Db 1854 GGTGGCCGTGAATGAGCCGAGGTGGCTTGTGTGAGAGAGAGAGAGTGTCTCTGGGGGG 1913
Qy 621 ProProGlyProLeuMetGlyGlnMetGlyGlyValArgArgGlyGlyArgGlyGlyProGly 640
Db 1914 CCTCTGAGCTTTATGGAACAGATGGGAGAGAGAGAGGCGGACGTGAGAGACCTGGG 1973
Qy 641 LysMetAspLysGlyGlnIleAspGlnGlyLysArgGlyLysArgGlyProTyr 656
Db 1974 AAAATGATTAAGGCGAGCACCTGTCAAGAACCAAGAGCCGCTTAC 2021

```

Search completed: February 21, 2005, 02:07:17  
 Job time : 7076.69 secs





XX	25-NOV-1993.
PD	
XX	19-MAY-1993; 93WO-FR000494.
FP	
XX	20-MAY-1992; 92FR-0006123.
FR	
XX	(CNRS ) CNRS CENT NAT RECH SCI.
PA	
XX	Aurias A, Delattre O, Desmaze C, Meloc T, Peter M, Plougastel B,
PI	Thomas G, Zucman J;
DR	WPI; 1993-386580/48.
DR	P-PSDB; AAR44555.
XX	
PT	New nucleic acid of EMS gene (and its hybrid(s) - contig. gene sequence
PT	involved in chromosomal trans-location, also derived mRNA, probes, fusion
FT	proteins etc., for diagnosis and treatment of Ewing sarcoma and melanoma.
XX	
PS	Disclosure; Fig 6; 123P; French.
XX	
CC	The probes 22RR3 and 22R12 were used to screen a human foetal brain cDNA
CC	library (Stratagene cat # 936206). The clone hRAOS was identified and
CC	sequenced. It represents the entire coding region and 3'-UTR of the Ews
CC	gene. (Updated on 25-MAR-2003 to correct PN field.)
XX	
SQ	Sequence 2371 BP; 639 A; 587 C; 659 G; 486 T; 0 U; 0 Other;
	Alignment Scores:
	Pred. No.: 1 69e-116. Length: 2371
	Score: 3633.00 Matches: 656
	Percent Similarity: 100.00% Conservative: 0
	Best Local Similarity: 100.00% Mismatches: 0
	Query Match: 100.00% Indels: 0
	DB: 2 Gaps: 0
US-10-791-017A-2 (1-656) x AAQ50643 (1-2371)	
OY	1 MetAlAsErThApTySeRtTYrSerGlnAlaAlaGlnGlnGlyTYrSerAla 20
Dd	25 ATGGCGTCCAGCATTAACATACTATAGCAGCTGCAGCGCAGCGGTACTACAGTGCT 84
OY	21 TyrThrAlaGlInPrOthrGlnGlyTYrAlaGlnTrpGlnAlaTYrGlyGlnInser 40
Dd	85 TACACGCCGCCAGCCACTCAAGGATATGACAGACCACCCAGGCAATTATGGCAACAAGC 144
OY	41 TYrGLyThrTYrGLyGlnPrOthrXzApValSerrTYrThrGlnAlaGlnThrThralaThr 60
Dd	145 TATGGAACCTATGACAGGCCCACTGATGTCAGCTATACCCAGGCTCAAGACCACTGCAAACC 204
OY	61 TyrGLyGlnThralaTYrAlaThrSerrTYrGLyGlnPrOthrGlnTYrThrPro 80
Dd	205 TATGGGACAGCCCGCTATGCAACTCTTATGACAGCCTCCACACTGGTTATACATCCA 264
OY	81 ThralaPrOglnAlaTYrSerGlnPrOvalGlnGlyTYrGLyThnGlyAlaTYrASpThr 100
Dd	265 ACTGCCCCCCAGGCAATACAGCACAGCTGTCCAGGGGTATGGCACTGTGCTTATGATACC 324
OY	101 ThrThralaThrValThrThrThrGlnAlaSerrTYrAlaAlaGlnSerrAlaTYrGLyThr 120
Dd	325 ACCACTGCTACAGTCAACCAACCCAGGCGTCTCTTATGACGCTCAGTCTGATATGGCACT 384
OY	121 GlnPrOlaTYrPrOlaTYrGLyGlnGlnPrOlaAlaAlaThralaPrOthrArgPrOglIn 140
Dd	385 CAACCTGCTTATCAACCTATAGGGCAGCGCAGCAGCACTGCACCTCAACAAGACCGCAG 444
OY	141 AspGIYAASnlArPrOthrGlnThrSerGlnPrOglInSerrThnGlyTYrASnGln 160
Dd	445 GATGGAACAAGCCCACTGAGCTAGTCAACTCACTACACAGGGGGGTATCAACAG 504
OY	161 ProSerLeuGLYTyrGLyGlnSerrAntTYrSerrTYrPrOglInValPrOglYSerrTYrPro 180
Dd	505 CCCAGGCTTAGTATATGACAGAATCACTATATCCCAGAGTACCTGGAGGACTACACCCC 564

QY	181	MecI	ProA	ProA	ThrAlaProP	ProSer	ProPro	ThrSer	ProSer	ThrGlnPro	200
Db	565	ATGAGCAG	CGATG	CACTG	CACTCACT	CACTCACT	CACTCACT	CACTCACT	CACTCACT	CACTCACT	624
QY	201	ThiSer	ProA	ProA	ProA	ProA	ProA	ProA	ProA	ProA	220
Db	625	ACTGTTAT	GTATG	TCAGAC	AGATTAC	TCACAGAA	CACTATGG	GCACCA	CCAGAC	AGCTAT	684
QY	221	GlyGln	GlnSer	ProGln	GlnSer	ProGln	GlnSer	ProGln	GlnSer	ProGln	240
Db	685	GGAACG	AGAGT	ACTTA	AGTCA	CAAGCA	AGCTATGG	GCACAG	CTTCCCA	CTAAGTAC	744
QY	241	ProPro	GlnThr	ProGln	ProSer	GlnAla	ProSer	GlnPro	ProSer	GlnSer	260
Db	745	CCACCCCA	ATCGGAT	CTCAAG	CAAGCTCA	AGTCA	TATTA	GCACCA	CAAGCA	GCAGCAGC	804
QY	261	TyrGly	GlnGln	ProSer	PheGln	GlnAla	ProSer	ProSer	ProSer	ProSer	280
Db	805	TACGGG	CAAGCA	AGATTAT	TCACAG	CAACCA	CCAGTAC	CAAGGCTT	ATATGG	GGAG	864
QY	281	GlnSer	GlnGly	PheSer	ProGln	GlnAla	ProSer	ProSer	ProSer	ProSer	300
Db	865	GAGCTG	AGAGAT	TTTTC	CCGAC	CAAGCA	AGATG	AGTGG	CTTCA	TACCGG	924
QY	301	GlyAla	GlnGly	PheAla	ProGly	GlyGln	ProSer	ProGly	GlyGln	ProGly	320
Db	925	GGCAGCG	GAAGAG	GGGAT	TTGAT	CTGAG	AGGCA	TAGAG	AGAGTGG	GGCGGAG	984
QY	321	ArgGly	GlnGly	ProSer	ProGln	GlnAla	ProSer	ProGln	GlnAla	ProSer	340
Db	985	CGCGGT	GAATGG	CAAGG	CTGAG	CAAGG	CTGCTT	CAATTA	AGCCTG	GTGAG	1044
QY	341	AspGln	GlyPro	AlaPhe	ProGln	ProGln	ProGln	ProGln	ProGln	ProGln	360
Db	1045	GATGA	AGCA	CCAGAT	CTTGA	CTGAG	CCCTCT	GTATG	CCAGAT	GAAGCA	1104
QY	361	SerAla	GlnGly	GlnGly	ProGln	ProGln	ProGln	ProGln	ProGln	ProGln	380
Db	1105	AGTCA	ATTAT	TATG	CAAGAT	TTAA	TAGAC	AGTGT	ACTTA	AGATCT	1164
QY	381	PheGln	GlnGly	GlnGly	ProGln	ProGln	ProGln	ProGln	ProGln	ProGln	400
Db	1165	TTTAA	CGAGT	GTGGG	CTTGT	TATTA	AGTAA	CAAGAA	CTGGG	CAACCAT	1224
QY	401	TyrLeu	AlaPhe	ProGln	ProGln	ProGln	ProGln	ProGln	ProGln	ProGln	420
Db	1225	TACCTG	CAAGCA	AGAA	CAAGAA	AGCCCA	AAAGG	CAATCC	CAAGG	CTCTTA	1284
QY	421	ProPro	AlaAla	AlaAla	ProGln	ProGln	ProGln	ProGln	ProGln	ProGln	440
Db	1285	CCCACT	CCAAAG	GTGCT	CGGAG	ATTTGAT	TGGAA	AGATTT	CAAGG	AGCAAACTT	1344
QY	441	LysAla	SerLeu	AlaAla	ProGln	ProGln	ProGln	ProGln	ProGln	ProGln	460
Db	1345	AAAGT	CTCCCT	GTGCT	CGGAG	AGGCT	CAATGA	AAAGTAT	CGGGG	GTCTG	1404
QY	461	ArgGln	GlyAla	ProGln	ProGln	ProGln	ProGln	ProGln	ProGln	ProGln	480
Db	1405	CGTAA	GGGCA	AGGCA	AGTCA	CCCACT	CGTGA	AGGCTCA	AGAGG	CCCAAGAG	1464
QY	481	GlyGly	ProGln	ProGln	ProGln	ProGln	ProGln	ProGln	ProGln	ProGln	500
Db	1465	GGGGA	AGCCCA	TGGT	CCCA	TGGAG	GGCTG	TGA	AGAA	TAGAG	1524
QY	501	GlyPro	AlaGly	ProGln	ProGln	ProGln	ProGln	ProGln	ProGln	ProGln	520
Db	1525	GGAC	CCCCGG	GGTTC	CCAG	AGAA	CCCTCT	GAG	AGAA	CCGTC	1584
QY	521	AspThr	GlnGly	ProGln	ProGln	ProGln	ProGln	ProGln	ProGln	ProGln	540
Db	1585	GACTG	GGAGT	GTCC	CAAT	CGGG	TTGG	AGAA	CCAGAA	CTTGG	1644

Qy	541	AsnGlnCysIySlyBAlAPProLyProGluGlyPheLeuProProPhaProProGly	560
Dy	1645	AACAGGTATAGGCCCCCAACCTCGAAGGCTTCCCTCCCGCACCTTCCCGCCCCGGGT	1704
Qy	561	GIyAPAPArgGlyIyArgGlyIyGlyProGlyIyMetArgGlyIyIyArgGlyIyGlyLeuMetAP	580
Dy	1705	GGTATCGTGCGACAGGTGGCCCTGGTGCAATGCGGGAGAGAGAGGTGGCTCATGAT	1764
Qy	581	ArgGlyIyGlyProGlyIyMetPheArgGlyIyIyArgGlyIyGlyIyAPAPArgGlyIyPheArg	600
Dy	1765	CGTGATGCTCCCGGGAATGTCAGAGGTGGCCCTGGTGAGAGACAGGTGGCTTCCT	1824
Qy	601	GIyGlyIyArgGlyIyMetAPAPArgGlyIyPheGlyIyGlyIyArgArgGlyIyGlyProGlyIy	620
Dy	1825	GGTGGCCGGGCGATGACCGAGGTGGCTTGGTGAGAGAGACAGGTGGCCCTGGGGGG	1884
Qy	621	ProProGlyIyProLeuMetGluGlnMetGlyIyIyArgArgGlyIyIyArgGlyIyGlyProGly	640
Dy	1885	CCCCCTGGACCTTGGATGGAACAGATGGAGAGAAAGAGAGACGTTGGAGGACTTGG	1944
Qy	641	LyMetAPAPLyGlyIyGlnIyIyArgGlnIyIyArgArgGlyIyArgAPAPArgProTy	656
Dy	1945	AAATGATTAAGGCGAGCACCGTCAGAGCGCAGAGATCGGCCCTAC	1992

RESULT 2

ADQ86032

ID ADQ86032 standard; cDNA, 2372 BP.

XX ADQ86032;

AC

XX

DT 07-OCT-2004 (first entry)

DE Human tumour-associated antigenic target (TAT) cDNA sequence #2904.

XX

XX human; tumour-associated antigenic target; TAT; cytostatic; gene therapy;

KM cancer; cell proliferative disorder; gene; ss.

XX

OS Homo sapiens.

XX

PN MO2004060270-A2.

XX

PD 22-JUL-2004.

XX

PF 15-OCT-2003; 2003WO-US029126.

XX

PR 18-OCT-2002; 2002US-0418988P.

XX

PA (GETH ) GENENTECH INC.

PA (WUTD/) WU T D.

PA (ZHOU/) ZHOU Y.

XX

PI Wu TD, Zhou Y;

XX

DR WPI; 2004-534300/51.

XX

PT New nucleic acid molecule and encoded polypeptide, for diagnosing,

XX preventing or treating cell proliferative disorders such as cancer.

XX

PS Claim 1; SEQ ID NO 2904; 5504bp; English.

CC

CC The present invention describes an isolated tumour-associated antigenic

CC target (TAT) nucleic acid comprising: (a) any of 4622 nucleotide

CC sequences (see SEQ ID NO:1 to 4622); (b) the full-length coding region of

CC (a); (c) the complement of (a) or (b); (d) a sequence that has 80%

CC sequence identity to (a)-(c); or (e) a sequence that hybridises to (a)-

CC (c). Also described: (1) an expression vector comprising the above

CC nucleic acid; (2) a host cell comprising the above expression vector; (3)

CC a process for producing a polypeptide; (4) an isolated polypeptide

CC comprising: (a) an amino acid sequence encoded by any of the above

CC nucleotide sequences; (b) an amino acid sequence encoded by the full-

CC length coding region of the above nucleotide sequences; or (c) a sequence

CC having at least 80% identical to (a) or (b); (5) a chimeric polypeptide

CC comprising the above polypeptide fused to a heterologous polypeptide; (6)

Seq	Sequence	2372 BP, 628 A, 589 C, 668 G, 487 T, 0 U, 0 Other
XX	Alignment Scores:	
XX	Pred. No.:	1,696-116 2372
XX	Score:	3633.00 656
XX	Percent Similarity:	100.00% 0
XX	Best Local Similarity:	100.00% 0
XX	Query Match:	100.00% 0
XX	DB:	13 Gaps: 0
US	US-10-791-017A-2 (1-656) x ADQ86032 (1-2372)	
QY	1 MetalaSerThrAspTyrSerThrTyrSerGlnAlaAlaGlnGlnGlyTyrSerAla	20
Db	44 ATGGGCTCCAGGATTTACAGTACCTATAGCCAAAGCTGACGAGCGAGGCTACAGTGGCT	103
QY	21 TyrThrAlaGlnProThrGlnGlyTyrAlaGlnThrThrGlnAlaGlyTyrGlnGlnSer	40
Db	104 TACACCGCCACGAGCCATCAAGGAATGACAGACACCCAGGCAATATGGCAACAAAGC	163
QY	41 TyrGlyThrTyrGlyGlnProThrAspValSerTyrThrGlnAlaGlnThrThrAlaThr	60
Db	164 TATGAGACCTTATGACAGCCCATATGTCACTTATCCAGGCTACAGACCACTGCAACC	223
QY	61 TyrGlyGlnThrAlaTyrAlaThrSerTyrGlyGlnProProThrGlyTyrThrThrPro	80
Db	224 TATGGGAGACCGGCTATGCAACTCTTATGACAGCCCTCCAGCTGTTATACTACCTCA	283
QY	81 ThrAlaProGlnAlaTyrSerGlnProValGlnGlyTyrGlyThrGlyAlaTyrAspThr	100
Db	284 ACTGCCCCCAGGACATACAGCAGCGCTGCCAGGGATATGGCACTGGTCTTATGATACC	343
QY	101 ThrThrAlaThrValThrThrThrGlnAlaSerTyrAlaAlaGlnSerAlaTyrGlyThr	120
Db	344 ACCACTGCTACAGTACCAACCAACGAGGCTCTTATGACGCTCAGCTGCAATATGGCACT	403
QY	121 GlnProAlaTyrProAlaTyrGlyGlnGlnProAlaAlaThrAlaProThrAspProGln	140
Db	404 CAGCGCTGTTATCCAGGCTATGSGGACGAGCCAGGACCACTGACCTTACAGACCGCAG	463
QY	141 AspGlyAsnLysProThrGlnThrSerGlnProGlnInserSerThrGlyGlyTyrAsnGln	160
Db	464 GATGGAACAAGACCCATGAGACTGATGCAACTCAATCTAGACAGGGGGTTATCAACACAG	523
QY	161 ProSerLeuGlyTyrGlyGlnSerAsnTyrSerTyrProGlnValProGlySerTyrPro	180
Db	524 CCGAGCTTGAATATGACAGAGTAACTTATCCCAAGTACTCTGGAGCTTACCCC	583

QY 181 MetGlnProValThrAlaProSerTyrProProthSerTyrSerSerThrGlnPro 200  
 Db 584 ATGAGCCAGTCACTGACCTCCATCTACCTCTACCACTATCTCTACCAAGCCG 643  
 QY 201 ThrSerTyrAspGlnSerSerTyrSerGlnGlnAsnThrTyrGlnProSerSerTyr 220  
 Db 644 ACTAGTTATGATCAGACCACTTACTCTCAGCAAGAACCTTATGGGCAACCGACAGCTAT 703  
 QY 221 GlnGlnGlnSerSerTyrGlnGlnGlnSerSerTyrGlnGlnGlnProProthSerTyr 240  
 Db 704 GGCAGCAGAGTGTGCTATGCTCAACAAGCAGCTATGGCAGCCTCCACTAGTTAC 763  
 QY 241 ProProGlnThrGlnSerTyrSerGlnAlaProSerGlnTyrSerGlnGlnSerSer 260  
 Db 764 CACCCCACTGGATCTTACAGCCCAAGCTCCAACTATACCAATACCAAGACAGCAGC 823  
 QY 261 TyrGlnGlnSerSerPheAspGlnAspHisProSerSerMetGlyValTyrGln 280  
 Db 824 TAGCGGAGCAGATTCTATCCAGCAAGACCAACCCAGTACATGGGTGTTATGGCAG 883  
 QY 281 GluSerGlnGlnPheSerGlnProGlnGlnAsnArgSerMetSerGlyProAspAsnArg 300  
 Db 884 GAGCTGAGAGATTTCTCCGAGCAGAGAACCCGAGCATGATGAGCTCCGATTAACCCG 943  
 QY 301 GlnArgGlnArgGlnPheAspArgGlnGlnMetSerArgGlnGlnArgGlnGln 320  
 Db 944 GGCAGGGGAAAGGGGATTTGATCGTAGAGCAGTAGAGAGAGGGGCGGAGAGGA 1003  
 QY 321 ArgGlnGlnMetGlnSerAlaGlnGlnArgGlnGlnPheAsnArgProGlnGlnProMet 340  
 Db 1004 CGCGGTGGAAATGGCAGCGCTGAGAGCAGAGTGGCTTCAATAAGCTGTGGAGCCCATG 1063  
 QY 341 AspGlnGlnProAspLeuAspLeuGlyProProValAspProAspGlnAspSerAspAsn 360  
 Db 1064 GATGAGAGCAGCAGATCTTGATCTAGAGCCCACTGTATCATGATGAGTGAAGCTGTGACAC 1123  
 QY 361 SerAlaIleTyrValGlnGlnGlnAspSerValThrLeuAspAspLeuAlaAspPhe 380  
 Db 1124 AGTGCATTTATGTCAGATTTAAATGACAGTGTGCTTAAATGATCTGGCAGACTTC 1183  
 QY 381 PheArgGlnGlnGlnValValGlnMetAsnLysArgThrGlnGlnProMetIleHisIle 400  
 Db 1184 TTTAAGCAGTGTGGGGTGTGTTAAGATGACAGAGAACTGGGCAACCATGATCCACATC 1243  
 QY 401 TyrLeuAspLysGlnThrGlnGlnProLysGlnAspAlaThrValSerTyrGlnAspPro 420  
 Db 1244 TACCTGAGCAAGAAAGAAAGAAAGCCCAAGGCGATCCACAGTCTCTATGAAGACCA 1303  
 QY 421 ProThrAlaLysAlaAlaValGlnTyrPheAspGlnTyrAspPheGlnGlnSerLysLeu 440  
 Db 1304 CCCACTGCCAAGGCTGCGTGGAAATGGTGTGATGGGAAAGATTTTCAAGGAGCAACTT 1363  
 QY 441 LysValSerLeuAlaArgLysLysProProMetAsnSerMetArgGlnGlnLeuProPro 460  
 Db 1364 AAGGTCTCCCTTGTGGAGAGAGCTCCATGAAGAGTATGGGGGTGTGCTGCCACCC 1423  
 QY 461 ArgGlnGlnArgGlnMetProProProLeuArgGlnGlnProGlnGlnProGlnGlnPro 480  
 Db 1424 CGTAGAGGAGAGGAGCAGCAGCAGCAGCTCGTAGAGGTCAGAGAGGCGCAGAGGCTCT 1483  
 QY 481 GlnGlnProMetGlnArgMetGlnGlnArgGlnGlnAspArgGlnGlnPheProProArg 500  
 Db 1484 GGGGAGCCCATGGGTCCATGGAGGAGGCGCTGAGAGAGATTAAGAGAGGCTTCTCCCAAGA 1543  
 QY 501 GlnProArgGlnSerArgGlnAspProSerGlnGlnGlnValGlnHisArgAlaGln 520  
 Db 1544 GAGACCCCGGGGTTCCGAGGGAACCTCTGTGAGAGAGAAAGTCCAGCAGCAGAGCTGGA 1603  
 QY 521 AspTyrGlnGlnProAspProGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 540  
 Db 1604 GACTGCGAGTGTCCCATCCGAGGTTGTGAAACCAAGCACTTCCGCTGAGAAACAGAGTGC 1663

QY 541 AsnGlnCysLysAlaProLysProGlnGlnPheLeuProProPheProProProGln 560  
 Db 1664 AACCACTGTAAGGCCCAAGCCTGAGAGCTTCTCTCCGCCACCTTCCGGCCCGGGGT 1723  
 QY 561 GlnAspArgGlnArgGlnGlnProGlnGlnMetArgGlnGlnArgGlnGlnLeuMetAsp 580  
 Db 1724 GGTGATCGTGGCAGAGGTGGCTGTGTGATGCGGGGAGAAAGAGTGGCTCCATGAGAT 1783  
 QY 581 ArgGlnGlnProGlnGlnMetPheArgGlnGlnArgGlnGlnAspArgGlnGlnPheArg 600  
 Db 1784 CGTGTGTCTCCGCTGAAATGTTCAAGAGTGGCCGTGTGAGAGACAGAGTGGCTTCCGT 1843  
 QY 601 GlnGlnArgGlnMetAspArgGlnGlnPheGlnGlnGlnGlnGlnGlnGlnGlnGln 620  
 Db 1844 GGTGGCCGGGCGATGAGCCGAGTGGCTTGTGTGAGAGAAAGAGAGTGGCTTGGGGGG 1903  
 QY 621 ProProGlnProLeuMetGlnGlnMetGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 640  
 Db 1904 CCCCCTGACCTTTGATGAGAACAGATGGAGAGAAAGAGAGAGAGAGAGAGAGAGAGAG 1963  
 QY 641 LysMetAspLysGlnGlnHisArgGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 656  
 Db 1964 AAATGATTAAGGCGAGCAGCTCAGAGCCGAGAGATCGGCTTAC 2011  
 RESULT 3  
 ABK84628  
 ID ABK84628 standard; cDNA; 2390 BP.  
 AC ABK84628;  
 XX  
 DT 14-AUG-2002 (first entry)  
 XX  
 DE Human cDNA differentially expressed in granulocytic cells #1199.  
 XX  
 KW Human; ss; granulocytic cell; DNA chip; bacterial infection;  
 KW viral infection; parasitic infection; protozoal infection;  
 KW fungal infection; sterile inflammatory disease; psoriasis;  
 KW rheumatoid arthritis; glomerulonephritis; asthma; chronic;  
 KW cardiac reperfusion injury; renal reperfusion injury; ARDS;  
 KW adult respiratory distress syndrome; inflammatory bowel disease;  
 KW Crohn's disease; ulcerative colitis; periodontal disease;  
 KW granulocyte activation; chronic inflammation; allergy.  
 XX  
 OS Homo sapiens.  
 XX  
 PN MO200228999-A2.  
 XX  
 PD 11-APR-2002.  
 XX  
 PF 03-OCT-2001, 2001MO-US030821.  
 XX  
 PR 03-OCT-2000, 2000US-0237189P.  
 XX  
 PA (GENE-) GENE LOGIC INC.  
 XX  
 PI Beazer-Barclay Y, Wetsman SM, Yamaga S, Vockley J;  
 XX  
 DR WPI; 2002-435328/46.  
 XX  
 PT Detecting granulocyte activation by detecting differential expression of  
 PT genes associated with granulocyte activation, which serves as diagnostic  
 PT markers that is useful for monitoring disease states and drug toxicity.  
 PS Claim 1, SEQ ID NO 1199, 114pp; English.  
 XX  
 CC The invention relates to detecting (M1) granulocyte (GC) activation  
 CC (GCA), by detecting the level of expression of gene(s) (Gs) identified by  
 CC DNA chip analysis as given in the specification, and comparing the  
 CC expression level to an expression level in an unactivated GC, where  
 CC differential expression of Gs is indicative of GCA. Also included are  
 CC modulating (M2) Gs by contacting GC with an agent that alters the  
 CC expression of at least one gene in Gs; (2) screening (M3) for an agent  
 CC capable of modulating GCA or an inflammation (especially chronic) in a



D6		1604 GACTGCGCAGTGTCCCAATTCGGGGTTTGTAACAACGAACCTGCCTGGAGAACAAGATGC 1663
OY		541 AengInCySlySAAlAProLYPProGlUgLyPhLeuPProPoProPhoProProGLy 560
D6		1664 AACCAAGTGTAAAGGCCCCCAAACCCTGMAAGGCTTCCCTCCGCCAACCTTCCGCCCCCCGGGT 1723
OY		561 GLVAPMPArRGVLVARSGLVGLYProglvGLYMecArgSLYLVARSGLYGLYLeuMeAsp 580
D6		1724 GGTAATCGTAGCACAGAGGTGGCCCTGGTGGCATTCGGGGAGAGAGAGGTGGCTCATGGAT 1783
OY		581 ArgSLylGLYProglvGLYMecPheARsgLvlYArgGLYLVAspARGLVLlYPheARg 600
D6		1784 CGTGTGGTGTCCCGGAGAAATTTCAGAGGTGGCCGTGTGGAGACAGAGGTGGCTTCGT 1843
OY		601 GLYLVLARGLVLMecASPARGLVLlYPheGLVLGLYLAARGALGLYLProglvGLY 620
D6		1844 GGTGGCGGGGCAATGACCGAGGTGGCTTTGTGTGAGGAACAAGAGGTGGCCCTGGGGGGG 1903
OY		621 ProProGLIYPRoleumMetGLUmEcGLVLGLYArgAVrgGLVLVARGLVLGLYProGLy 640
D6		1904 CCCCCTGACCTTTGATGTGAACAGATGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1963
OY		641 LysMeCAsPLYSGLVLUMHleARGGLngLUARGARVASPARgProTYr 656
D6		1964 AAAATGATTAAGCGGACGCCGTGAGAGCGCAGACGATCGGCCTTAC 2011
<hr/>		
RESULT 4		
ABK64822	ID	ABK64822 standard; DNA, 2390 BP.
XX	AC	ABK64822;
XX	DT	18-JUN-2002 (first entry)
DE	XX	Human benign prostatic hyperplasia gene #717.
XX	KM	Human; benign prostatic hyperplasia; BPH; prostate cancer; gene; de.
OS	XX	Homo sapiens.
PX	PN	MO200212440-A2.
PD	XX	14-FEB-2002.
XP	PP	07-AUG-2001; 2001WO-US024708.
XX	PR	07-AUG-2000; 2000US--0223323P.
PR	XX	05-JUN-2001; 2001US--00873319.
XA	PA	(GENE-) GENE LOGIC INC.
PA	PA	(NISR) JAPAN TOBACCO INC.
PI	PI	Munger WE, Kulkarni P, Getzenberg RH, Waga I, Yamamoto J;
XX	DR	WPI, 2002-257476/30.
XX	XX	
PS	XX	Disclousure, Page 399-400; 444pp; English.
CC	CC	The invention relates to a method of diagnosing (I) the onset or progression of benign prostatic hyperplasia (BPH), or screening (II) for or identifying an agent that modulates the onset or progression of BPH. The method is based on changes in gene expression in BPH tissue isolated from patients exhibiting different clinical states of prostate hyperplasia as compared to normal prostate tissue. (I) comprises detecting the expression levels of one or more genes in prostate cells from the subject that are differentially regulated compared to normal prostate cells. (II) comprises preparing a first gene expression profile of BPH cells or BPH-like cell population, exposing the cells to the

CC agent, preparing a second gene expression profile of the agent exposed  
CC cells, and comparing the first and second gene expression profiles. (I)  
CC is useful for diagnosing the onset or progression of BPH. (II) is useful  
CC for identifying an agent that modulates the onset or progression of BPH..  
CC The methods are useful to present information identifying the expression  
CC level in a tissue or cells, by comparing the expression level of genes  
CC given in the specification in the tissue or cells to the level of  
CC expression of gene in the database, and displaying the expression levels  
CC of at least one gene in the tissue or cell sample compared to the  
CC expression level in BPH. Agents using (II) are useful for treating BPH or  
CC prostate cancer. ABK64106-ABK64860 represent human benign prostatic  
CC hyperplasia gene sequences of the invention

**SQ** Sequence 2390 BP; 645 A; 589 C; 668 G; 488 T; 0 U; 0 Other;

**Alignment Scores:**

Score:	3633.00	Matches:	656
Concentration:	100.00%	Concentration:	0

Best Local Similarity:	100.00%	Mismatches:	0
Custom Watch:	100.00%	Traps:	0

DB:	6	Groups:	0
-----	---	---------	---

US-10-791-017A-2 (1-656) x ABK64822 (1-2390)

QY 1 MetAlaSerThrAspTyrSerThrTyrSerGlnAlaAlaGlnGlnGlyTyrSerAla 20

Db 4 ATGGCGTCCACGATTACCTATAGCCCAAGCTGCAGCGCAGCAGGGCTACAGTGT 10

21 TyrThrAlaGlnProThrGlnGlyTyrAlaGlnThrThrGlnAlaTyrGlyGlnGlnSer 400

Db 104 TACACCGCCAGCCCACTCAAGGATATGCACAGACCACCAGGCATATGGGCAACAAAGC 16

41 TyrglyThryrglInProThraspvalSertyrThrglnAlaglnThrThralaThr 60

Db 164 TATGGAACCTATGGACAGCCCACTGATGTCACTATATACCGAGGCTCAGACCACCTGCAACC 22

61 TyrgIyGlnThrAlaTyraIatThrSerTyrgIyGlnProProThrgIyTThrThrPro 80

Db 224 TATGGCAGACCGCCTATGCACTTCTTATGGACAGCCTCCCACTGGTTATACTACTCCA 28

81 ThrAlaProGlnAlaTyrSerGlnProValGlnGlyTyrGlyThrGlyAlaTyrAspThr 100

Db 284 ACTGCCCCCAGGCATACAGCCAGCCTGTCTCAGGGGTATGGCACTGGTCTTATGATACC 34

101 ThrThrAlaThrValThrThrThrglnAlaSerTyrAlaAlaGlnSerAlaTyrGlyThr 12

Db 344 ACCACTGCTACGTCAACCAACCAAGGCTCCTATGCAAGCTCAGTCTGCATATGGCACT 400

QY 121 GlnProAlaTyrProAlaTyrGlnGlnProAlaAlaThrAlaProThrArgProGln 14

Db 404 CAGCCTGCTTATCCAGCCTATGGGCAGCAGCCAGCAGCCACTGCACTTACAGACCCGAG 40

141 ASPGLYASNLVSPROTHRGILUTHRSERGINPROGLINSESETHRGILGYLYRASNGIN 16

Db 464 GATGGAACAGCCCACTGAGACTAGTCAACCTCAATCTAGCACAGGGGTTACAAACAG 52

161 ProSerLeuGLY TyrGLY GlnSerAsn TyrSer TyrProGlnValProGLY SerTyrPro 166

Db 524 CCACGCTAGGATATGGACAGAGTAACTACAGTATATCCCCAGGTACCTGGAGCTACCCC 528

181 MetGlnProValThrAlaProProSerTyrProProThrSerTyrSerSerThrGlnPro 200

Db 584 ATGCAGCCAGTCACTGCACCTCCATCCTACCCCTCCCTACCCAGCTATTTCTCTACACAGCCG 64

201 ThrSerTyrAspGlnSerSerTyrSerGlnGlnAsnThrTyrGlnGlnProSerSerTyr 22

Db 644 ACTAGTTATGATCAGAGCAGTTACTCTCAGCAGAACCACTATGGGCAACCGAGCAGCTAT /

221 GYGLNGNSERTYRGLYGLNGNSERTYRGLYGLNGNPROPROINRSEYF 24

Db 704 GGACAGCAGTAGCTATGTCACCAAGCAGCTATGGCAGCAGCCCTCCACCTAGTAC /



QY 241 ProProGlnThrGlySerTyrSerGlnAlaProSerGlnTyrSerGlnIleSerSerSer 260  
 DB 764 CACCCCAAACTGGATCTCAAGCCAAAGCTCCAAATGATCCAAAGAGAGAGAGC 823  
 QY 261 TyrGlyGlnIleSerSerPheArgGlnAspHisProSerSerMetGlyValTyrGlyGln 280  
 DB 824 TAGGGGAGAGAGATTGATTCAGACAGAGACCAACCCAGTACATGGGTGTTTATGGGAG 883  
 QY 281 GluSerGlyGlyPheSerGlyProGlyGlyAsnArgSerSerSerGlyProAspAsnArg 300  
 DB 884 GAGTCTGGAGGATTTCTCCGACAGAGAGAAACCGAGCATGATGGCCCTGATTAACGG 943  
 QY 301 GlyArgGlyArgGlyGlyPheAspArgGlyGlyMetSerArgGlyGlyArgGlyGlyGly 320  
 DB 944 GGCAGGGGAAAGAGGGGATTGATCTGAGAGCATGAGAGAGTGGCGGGAGAGAGAG 1003  
 QY 321 ArgGlyGlyMetGlySerAlaGlyGlyGlyGlyGlyGlyPheAsnArgProGlyGlyProMet 340  
 DB 1004 CGCGGTGGATGGGACGCTGAGAGAGAGTGGCTTCAATAGACCTGGTGGACCAATG 1063  
 QY 341 AspGlyGlyProAspLeuAspLeuGlyProProValAspProAspGlyAspSerAspAsn 360  
 DB 1064 GATGAAGAGACCATCTTGATCTGAGCCCTCTGTATGATCCAGATGAAGACTGACAC 1123  
 QY 361 SerAlaIleTyrValGlnGlyLeuAspSerValThrLeuAspAspLeuAlaAspPhe 380  
 DB 1124 AGTGCAATTTATGACAGAGATTAAATGACAGTGTGCTCAGATGATCTGGACACTTC 1183  
 QY 381 PheLeuGlnCysGlyValValIleMetAsnIleArgThrGlyGlnProMetIleHisIle 400  
 DB 1184 TTTAAGAGTGGGTGTTTGAATGAACAGAGAACTGGGACCAACCAATGATCCATC 1243  
 QY 401 TyrLeuAspIleGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 420  
 DB 1244 TACCTGGACAAAGAGAAAGAGAAAGCCCAAGGCGCATCCACAGTCTCTATGAAGACCA 1303  
 QY 421 ProThrAlaIleValAlaValGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 440  
 DB 1304 CCACATGCGCAAGGCTGCGTGGATGGTGTGATGGAGAAATTTTCAAGGAGAGAACTT 1363  
 QY 441 LysValSerLeuAlaArgGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 460  
 DB 1364 AAAGTCTCTCTTCTCGAGAGAGAGAGCTCCATGACAGTATCGGGGTGTTCTCCACC 1423  
 QY 461 ArgGlyGlyArgGlyMetProProProLeuArgGlyGlyGlyProGlyGlyPro 480  
 DB 1424 CGTAGGAGGAGAGGATGACCAACCACTCCGTGAGAGTCCAGAGGCGCCAGAGGTCCT 1483  
 QY 481 GlyGlyProMetGlyArgMetGlyGlyArgGlyGlyArgGlyGlyPheProProArg 500  
 DB 1484 GGGGAGACCCATGGGTGCGCATGGAGGCGCGTGGAGAGATGAGAGAGGCTTCCCTCCAA 1543  
 QY 501 GlyProArgGlySerArgGlyAsnProSerGlyGlyGlyValGlnHisIleArgAlaGly 520  
 DB 1544 GAGACCCCGGGGTTCCCGAGGAGAACCTCTCGAGAGAGAGAAAGTCCAGCAGCAAGCTG 1603  
 QY 521 AspTyrGlnCysAspProAsnProGlyGlyGlyAsnGlnAsnAspAlaTyrThrGlnGly 540  
 DB 1604 GACTGGCAGTGTCCCATCCGGGTGGGAAACCAAGAACTTCCGTGAGAAACAGAGTGC 1663  
 QY 541 AsnGlyCysIleValAspProGlyGlyPheLeuProProPheProProProGly 560  
 DB 1664 AACCAAGTGAAGGCCCAAGAGCTGAAAGGCTTCTCCGCGCACCTTTCGCGCCCGG 1723  
 QY 561 GlyAspArgGlyArgGlyGlyProGlyGlyMetArgGlyGlyArgGlyGlyLeuMetAsp 580  
 DB 1724 GGTGATGTGGCAGAGGTGCGCTGTGGCATGCGGGGAGAGAAAGTGGCTCATGAT 1783  
 QY 581 ArgGlyGlyProGlyGlyMetPheArgGlyGlyArgGlyGlyAspArgGlyGlyPheArg 600  
 DB 1784 CGTGTGTGTCCCGTGGAAATGTTCAAGAGTGGCGGTGTGAGAGCAAGAGTGGCTTCGT 1843  
 QY 601 GlyGlyArgGlyMetAspArgGlyGlyPheGlyGlyGlyArgArgGlyGlyProGlyGly 620

DB 1844 GGTGGCCGGGCAATGAGCCAGAGTGGCTTGTGGAGAGAAAGAGAGTGGCTTGGGGGG 1903  
 QY 621 ProProGlyProLeuMetGlnIleMetGlyGlyArgArgGlyGlyArgGlyGlyProGly 640  
 DB 1904 CCCCCTGAGACTTTGATGAGAAAGATGGAGAGAAAGAGAGAGAGAGAGAGAGAGAGAG 1963  
 QY 641 LysMetAspIleGlyGlyGlnHisArgGlnGlyArgArgAspArgProTyr 656  
 DB 1964 AAATGGATTAAGCGCAGCACCGTCAAGAGAGAGAGAGATGCGCCCTAC 2011  
 RESULT 5  
 ID ABN97274 standard; DNA; 2390 BP.  
 AC ABN97274;  
 DT 13-AUG-2002 (first entry)  
 DE Gene #3772 used to diagnose liver cancer.  
 XX Gene; liver cancer; de; hepatocellular carcinoma; hepatotropic;  
 XX metastatic liver tumour; cytostatic; expression profile; disease state;  
 XX disease progression; drug toxicity; drug efficacy; drug metabolism.  
 OS Homo sapiens.  
 XX  
 PN WO200229103-A2.  
 PD 11-APR-2002.  
 PF 02-OCT-2001; 2001MO-US030589.  
 XX  
 PR 02-OCT-2000; 2000US-0237054P.  
 XX  
 PA (GENE-) GENE LOGIC INC.  
 XX  
 PI Horne D, Alvares C, Peres-Da-Silva S, Vockley JG;  
 XX WPI; 2002-426119/45.  
 DR  
 XX  
 PT Diagnosing and detecting the progression of liver cancer, hepatocellular  
 PT carcinoma or metastatic liver tumor in a patient, involves detecting the  
 PT level of expression of two or more genes in a liver tissue sample.  
 XX  
 PS Claim 1; SEQ ID NO 3772; 298bp; English.  
 XX  
 CC The invention relates to a novel method for diagnosing and detecting the  
 CC progression of liver cancer, hepatocellular carcinoma or metastatic liver  
 CC tumour in a patient, and differentiating metastatic liver cancer from  
 CC hepatocellular carcinoma in a patient, involving detecting the level of  
 CC expression of two or more genes represented in ABN93503-ABN97455 in a  
 CC tissue sample. The method of the invention has hepatotropic, and  
 CC cytostatic activity. The method is useful for diagnosing and detecting  
 CC the progression of liver cancer, hepatocellular carcinoma and metastatic  
 CC liver carcinoma in a patient. The method is useful for identifying  
 CC expression profiles which serve as useful diagnostic markers as well as  
 CC markers that can be used to monitor disease states, disease progression,  
 CC drug toxicity, drug efficacy and drug metabolism. Note: The sequence data  
 CC for this patent did not form part of the printed specification, but was  
 CC obtained in electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pat\_sequences  
 CC  
 SQ Sequence 2390 BP; 645 A; 589 C; 668 G; 488 T; 0 U; 0 Other;  
 XX  
 Alignment Scores:  
 Pred. No.: 1,7e-116 Length: 2390  
 Score: 3633.00 Matches: 656  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: 6 Gaps: 0

US-10-791-017A-2 (1-656) X ABN97274 (1-2390)

QY 1 MetAlaSerThrAspTyrSerThrTyrSerGlnAlaAlaAlaGlnGlnGlnTyrTyrSerAla 20  
 Db 44 ATGGCGGTCACGAGATTACATTAAGCTTAAGCCAGCTGACGCGACGCGGGCTCAAGAGCT 103  
 QY 21 TyrThrAlaGlnProThrGlnGlnTyrAlaGlnThrThrGlnAlaTyrGlnGlnGlnSer 40  
 Db 104 TACACCGCCGACGACCTCAAGATATGACAGACCCACGCGCATATGGGCAACAAAGC 163  
 QY 41 TyrGlnTyrThrTyrGlnGlnProThrAspValSerTyrThrGlnAlaGlnThrThrAlaThr 60  
 Db 164 TATGAAACCTATGGACAGCCCACTGATGTCACTATCCAGGGCTCAGACCACTGCAACC 223  
 QY 61 TyrGlnGlnThrAlaTyrAlaThrSerTyrGlnGlnProThrGlnTyrThrThrPro 80  
 Db 224 TATGGGAGACCGCTATGCACTTCTTATGAGACAGCTCCCACTGGTGTATCTACTCCA 283  
 QY 81 ThrAlaProGlnAlaTyrSerGlnProValGlnGlnTyrGlnTyrThrGlnAlaTyrAspThr 100  
 Db 284 ACTGCCCCGACGATACAGCCAGCTGTCCAGGGGATATGGCACTGGTCTTATGATACC 343  
 QY 101 ThrThrAlaThrValThrThrThrGlnAlaSerTyrAlaAlaGlnSerAlaTyrGlnTyr 120  
 Db 344 ACCACTGCTACAGTCAACACCAACCGCTCTTATGACCTCAGTCTGATATGGCACT 403  
 QY 121 GlnProAlaTyrProAlaTyrGlnGlnGlnProAlaAlaThrAlaProThrArgProGln 140  
 Db 404 CAGCTGCTTATCCAGCTATGGGACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 463  
 QY 141 AspGlnValAsnLysProThrGlnThrSerGlnProGlnSerSerThrGlnTyrTyrAsnGln 160  
 Db 464 GATGGAAACAAAGCCCACTGAGACCTAGTCAACCTCAATGACACAGGGGCTTACAAACAG 523  
 QY 161 ProSerLeuGlnTyrGlnGlnSerAsnTyrSerTyrProGlnValProGlnSerTyrPro 180  
 Db 524 CCACGCTTATGAGATATGACAGAGTAACTAAGTATCCACAGGATCCAGGAGCTCAACCC 583  
 QY 181 MetGlnProValThrAlaProProSerTyrProProThrSerTyrSerSerThrGlnPro 200  
 Db 584 ATGAGCGACAGTCACTGACCTCCACTACCTCTCAACAGTATCTCTTACACAGCCG 643  
 QY 201 ThrSerTyrAspGlnSerSerTyrSerGlnGlnAlaSerThrTyrGlnGlnProSerSerTyr 220  
 Db 644 ACTAGTATGATCAAGACAGTACTCTCAAGAAACCTATATGGGACCAACGACAGCTAT 703  
 QY 221 GlnGlnGlnSerSerTyrGlnGlnGlnSerSerTyrGlnGlnGlnProProThrSerTyr 240  
 Db 704 GGAACAGACAGTATGCTATGCTCAACAAAGCAGCTATATGGGACAGCTCCCACTAGTTAC 763  
 QY 241 ProProGlnThrGlnSerTyrSerGlnAlaProSerGlnThrSerGlnGlnSerSerSer 260  
 Db 764 CCAACCCCAACTGATCTTACAGCCAAAGCTCAAGTAAATATAGCCAAACAGACAGCAGC 823  
 QY 261 TyrGlnGlnGlnSerSerPheArgGlnAspHisProSerSerMetGlnValTyrGlnGln 280  
 Db 824 TAGGGGACAGAGATTCACTCCAGACAGACCAACCCAGTACATGGGTGTTATGGGACG 883  
 QY 281 GlnSerGlnTyrPheSerGlnTyrProGlnGlnAlaAsnArgSerMetSerGlnTyrProAspAsnArg 300  
 Db 884 GAGCTGTGAGAGATTTCCTCGGACCGAGAGAACCGGAGCATATGATGCTCCGTGATACCGG 943  
 QY 301 GlnTyrGlnTyrArgGlnTyrPheAspArgGlnGlnTyrMetSerArgGlnTyrArgGlnTyr 320  
 Db 944 GGCAGGGGAAAGAGGGGATTTGATCGTGAAGCATATGACAGAGTGGGCGGGGAGAGGGA 1003  
 QY 321 ArgGlnTyrMetGlnSerAlaGlnTyrGlnTyrGlnTyrPheAsnLysProGlnTyrProMet 340  
 Db 1004 CGCGGTGGAATGGGCAACCGCTGAGAGACAGAGTGGCTTCAATATAGCTGTGATGCCATG 1063  
 QY 341 AspGlnGlnTyrProAspLeuAspLeuGlnTyrProProValAspProAspGlnAspSerAspAsn 360  
 Db 1064 GATGAAGGACCAAGATCTTGAATCTATGGCCCTCTGTATATCATGATGAAGACTCTGACAAAC 1123

QY 361 SerAlaIleTyrValGlnGlnTyrLeuAsnAspSerValThrLeuAspAspLeuAlaAspPhe 380  
 Db 1124 AGTGCATATTTATGATCAAGAGATTAAATGACAGTGTACCTTATGATATCTGGACAGCTTC 1183  
 QY 381 PheTyrGlnCysGlnValValTyrMetAsnLysArgThrGlnGlnProMetIleHisGln 400  
 Db 1184 TTAAACAGATGGGTGTTGTTAAGATGAACAAGAACTGGGCAACCATGATCCATCATC 1243  
 QY 401 TyrLeuAspLysGlnThrGlnTyrProLysGlnAspAlaThrValSerTyrGlnAspPro 420  
 Db 1244 TACCTGACACAGAAACAGAAAGAGCCCAAGGGGATGCCACAGTGTCTTATGAAGACCCA 1303  
 QY 421 ProThrAlaLysAlaAlaValGlnTyrPheAspGlnTyrAspAspPheGlnGlnSerLysLeu 440  
 Db 1304 CCACCTGCAAGGCTCTCCGTGATATGTTGATGGAAAGATTTTCAAGGACCAAACTT 1363  
 QY 441 LysValSerLeuAlaArgLysLysProProMetAsnSerMetArgGlnTyrLeuProPro 460  
 Db 1364 AAAGTCTCCCTTGTCTGGAAAGACCTCCAAATGAACAGTATGGGGGTGTGTGCCACCC 1423  
 QY 461 ArgGlnGlnTyrGlnTyrMetProProProLysArgGlnTyrProGlnTyrProGlnTyrPro 480  
 Db 1424 CGTGAAGGACAGGACATGCAACCACTCCGTGAGGTCCAGAGAGCCCAAGAGATGCTT 1483  
 QY 481 GlnTyrProMetGlnTyrMetGlnTyrArgGlnTyrAspArgGlnTyrLysPheProArg 500  
 Db 1484 GGGGAGCCCATGGGTGTGATGGAGGCTCGTGAAGAGATAGAGAGAGCTTCCCTCCAAAG 1543  
 QY 501 GlnProArgGlnSerArgLysAsnProSerGlnTyrGlnLysValGlnHisArgAlaGln 520  
 Db 1544 GAGCCCGGGGTTCCGAGGGAACCCCTCTGAGGAGGAAACGTCCAGCAGCCAGCTGGA 1603  
 QY 521 AspTyrGlnCysPheAsnProGlnTyrGlnLysAsnGlnAsnPheAlaTTPATGTTGluCys 540  
 Db 1604 GACTGCAAGTGTCCCAATCCGGGTGTGGAACCAAACTTGCCCTGAGAAACAGAGTGC 1663  
 QY 541 AsnGlnCysLysAlaProLysProGlnGlnTyrPheLeuProProProPheProProGln 560  
 Db 1664 AACCATGTAAAGGCCCCCAAGCTGAAGCTTCTCTCCGACCTTCCGCCCCGGGT 1723  
 QY 561 GlnAspArgGlnTyrArgGlnTyrProGlnTyrMetArgGlnTyrArgGlnTyrLeuMetAsp 580  
 Db 1724 GGTGATCGTGGCAGAGTGGCTTGTGGCATTCGGGGAGAAAGTGGCTCATGAT 1783  
 QY 581 ArgGlnTyrProGlnTyrMetPheArgGlnTyrArgGlnTyrLysAspArgGlnTyrPheArg 600  
 Db 1784 CGTGTGTGTCCTCGGTGAATGTTCAAGAGTGGCCGTGTGGAGACAGAGTGGCTTCCGT 1843  
 QY 601 GlnTyrArgGlnTyrMetAspArgGlnTyrPheGlnTyrGlnTyrArgArgGlnTyrProGlnTyr 620  
 Db 1844 GGTGGCCGGGCAATGAAACGAGGTGGCTTGTGTGAGAGAAAGAGAGTGGCTTGGGGGG 1903  
 QY 621 ProProGlnTyrProLeuMetGlnGlnMetGlnTyrArgArgGlnTyrArgGlnTyrProGln 640  
 Db 1904 CCCCTGGAACCTTATATGAAACAGATGGGAGAAAGAGAGAGAGAGAGAGAGAGAGAGAG 1963  
 QY 641 LysMetAspLysGlnTyrHisArgGlnGlnTyrArgArgAspArgProTyr 656  
 Db 1964 AAATATGATTAAGGACAGACCGTCAAGAGCCGAGAGATCGCCCTTAC 2011

RESULT 6  
 ADS16277  
 ID ADS16277 standard; cDNA, 2390 BP.  
 XX  
 AC ADS16277;  
 XX  
 DT 18-NOV-2004 (first entry)  
 XX  
 DE Human cDNA encoding the Ewing sarcoma protein Segid 1.  
 XX human; ss; gene; Ewing sarcoma; EWS; prostatic cancer; alopecia; acne;  
 KW hypogonadism; androgen-resistance syndrome; testicular feminization.

XX OS Homo sapiens.  
 XX FH Key Location/Qualifiers  
 FT CDS 44..2014  
 FT /\*tag= a  
 FT /product= "Ewing sarcoma protein"  
 XX BN1455190-A1.  
 XX PD 08-SEP-2004.  
 XX PF 16-FEB-2004; 2004EP-00003422.  
 XX PR 04-MAR-2003; 2003DE-01009280.  
 XX PR 25-APR-2003; 2003US-0465652P.  
 XX (SCHD ) SCHERING AG.  
 XX Obendorf M, Wolf S;  
 XX WPI; 2004-627861/61.  
 XX P-PSDB; ADS16278.  
 XX Determining the hormonal effects of substances, used to identify  
 PT pharmaceuticals, e.g. for treatment of androgen receptor dysfunction,  
 PT from modulating interaction between nuclear receptors and Ewing sarcoma  
 PT protein.  
 XX Claim 7; SEQ ID NO 1; 30pp; German.  
 XX This invention relates to a novel modulators that alter the interaction  
 CC between the Ewing sarcoma protein (EWS) and its nuclear receptor, as well  
 CC as the screening method thereof. Specifically, it refers to determining  
 CC and identifying a hormonal effect brought about by test compounds that  
 CC modulate either the binding of EMS to the nuclear receptor or the ligand-  
 CC induced activity of this receptor. The present invention describes the  
 CC nuclear receptors as including oestrogen, progesterone, thyroid hormone,  
 CC vitamin D, and retinoic acid receptors, most preferably they are androgen  
 CC receptors. Accordingly, these modulators may be used in the development  
 CC of pharmaceutical compositions that can diagnose and be used to treat  
 CC diseases associated with receptor dysfunction such as prostatic cancer,  
 CC alopecia, acne, hypogonadism and androgen-resistance syndrome e.g.  
 CC testicular feminization. This method provides reliable, sensitive, e.g.  
 CC simple, inexpensive and rapid assessment of the hormonal effects of these  
 CC test compounds. This polynucleotide sequence is the cDNA encoding the  
 CC human Ewing sarcoma protein of the invention.  
 XX  
 XX SQ Sequence 2390 BP; 646 A; 589 C; 668 G; 487 T; 0 U; 0 Other;  
 Alignment Scores:  
 Pred. No.: 1,7e-116 Length: 2390  
 Score: 3633.00 Matches: 656  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: 13 Gaps: 0  
 US-10-791-017a-2 (1-656) x ADS16277 (1-2390)  
 QY 1 MetAlaSerThrAspTyrSerThrTyrSerGlnAlaAlaGlnGlnGlyTyrSerAla 20  
 DB |||||  
 DB 44 ATGGCGCTCCACGGATTACAGTACTTATAGCAAGCTGACGCGCAGCGGGCTACAGTGC 103  
 QY 21 TTTThAlaGlnProThrGlnGlyTyrAlaGlnThrThrGlnAlaTyrGlyGlnGlnSer 40  
 DB |||||  
 DB 104 TACACGGCCCAAGCCCACTCAAGATATGCACACCAACGCAATATGGGCAACAAAGC 163  
 QY 41 TTTGlyThrTyrGlyGlnProThrAspValSerTyrThrGlnAlaGlnThrThralaThr 60  
 DB |||||  
 DB 164 TATGGAACTATAGACAGCCCACTGATGTACGATATACCAAGCTCAGACCACTGCAACC 223  
 QY 61 TTTGlyGlnThrAlaTyrAlaThrSerTyrGlyGlnProThrGlnGlyTyrThrThrPro 80

DB 224 TATGGGAGACCCCTCATGCACTTCTTATGACAGCTCCCACTGGTTTACTACTCCA 283  
 QY 81 ThrAlaProGlnAlaTyrSerGlnProValGlnGlyTyrGlyAlaTyrAspThr 100  
 DB ACTGCCCCCAAGGACATACAGCCGCTGTCAGAGGGGATAGGCACTGGTGTCTTATGATACC 343  
 QY 101 ThrThrAlaThrValThrThrThrGlnAlaSerTyrAlaAlaGlnSerAlaTyrGlyThr 120  
 DB 344 ACCACTGTACATACACCAACACAGCCGCTCTTATGACAGCTCACTGATATGACACT 403  
 QY 121 GlnProAlaTyrProAlaTyrGlyGlnGlnProAlaAlaThrAlaProThrAspProGln 140  
 DB 404 CAGCTGCTTATCAAGCTTATGGGACAGCCCAACGCACTCACTCAACAGACGCGAG 463  
 QY 141 AspGlyAsnLysProThrGlnThrSerGlnProGlnSerSerThrGlyTyrAsnGln 160  
 DB 464 GATGAAACAGCCCACTGACCTAGTCAACCTCATCTTACAGACAGGGGGTTTCAACAG 523  
 QY 161 ProSerLeuGlyTyrGlyGlnSerAsnTyrSerTyrProGlnValProGlySerTyrPro 180  
 DB 524 CCCAGCCTAGATATGACAGAGTAACTTACATTATCCCAAGGTTACCTGGAGCTACCCC 583  
 QY 181 MetGlnProValThrAlaProProSerTyrProProThrSerTyrSerSerThrGlnPro 200  
 DB 584 ATGACGCGAGTCACTGACCTCATCTTACCTTCCCTTACAGAGCTATTTCTTACAGCCG 643  
 QY 201 ThrSerTyrAspLysSerTyrSerGlnGlnAsnThrTyrGlyGlnProSerSerTyr 220  
 DB 644 ACTAGTTATGATACAGAGCACTTACTCTCAGAGAAACCTTATGGGCAACGAGCACTAT 703  
 QY 221 GlyGlnGlnSerSerTyrGlyGlnGlnSerSerTyrGlyGlnGlnProThrSerTyr 240  
 DB 704 GGCACACAGAGTGTCTATGCTCAACAAACAGCTATGGGACAGCACTCCCACTAGTTAC 763  
 QY 241 ProProGlnThrGlySerTyrSerGlnAlaProSerGlnTyrSerGlnGlnSerSer 260  
 DB 764 CACACCCAACTGATCTTACAGCAAGCTCCAACTCAATATATGCAACAGACAGCAACAGC 823  
 QY 261 TTTGlyGlnGlnSerSerPheArgGlnAspHisProSerSerMetGlyValTyrGlyGln 280  
 DB 824 TACGGGACAGAGTTTCACTTCCAGACAGACCAACCAAGTATGATGGTGTATTATGGGAG 883  
 QY 281 GluSerGlyGlyPheSerGlyProGlyGluAsnArgSerMetSerGlyProAspAsnArg 300  
 DB 884 GAGTCTGAGAGTATTTCCGAGACAGAGAAACCGAGCATGATGAGCTGCTTAACCGG 943  
 QY 301 GlyArgGlyArgGlyGlyPheAspArgGlyGlyMetSerArgGlyGlyArgGlyGly 320  
 DB 944 GGCAGGGGAAAGGGGGGATTTGATCGTGAAGCATGACAGAGTGGCGGGAGAGAGCA 1003  
 QY 321 ArgGlyGlyMetGlySerAlaGlyGluArgGlyGlyPheAsnLysProGlyGlyPromet 340  
 DB 1004 CCGCGTGAATGGGACGCGCTGAGAGCGAGGTGGCTTCAATAGCTGTGTGAGACCCAG 1063  
 QY 341 AspGlyGlyProAspLeuAspLeuGlyProProValAspProAspGlyLysAspAsn 360  
 DB 1064 GATGAAGGACAGATCTTATCTAGGCCCACTGTATGATCCAGATGACATCTGCAAC 1123  
 QY 361 SerAlaIleTyrValGlnGlnLysLeuAsnAspSerValThrLeuAspAspLeuAlaAsp 380  
 DB 1124 AGTGCATTTATGTACAGAGATTAATGACATGTGACCTTATGATGTCTGACAGACTTC 1183  
 QY 381 PheLysGlnCysGlyValValLysMetAsnLysArgThrGlyGlnProMetIleHisIle 400  
 DB 1184 TTTTACAGTGTGGGTTTATTAAGTAAACAGAGAACTGGGCAACCACTATTCACATC 1243  
 QY 401 TTTLeuAspLysGlyThrGlyLysProLysGlyAspAlaThrValSerTyrGlyLysPro 420  
 DB 1244 TACCTGGAACAAGAAACAGAAAGCCCAAGGCGATGACAGAGTGTCTTATGAGAGCCCA 1303  
 QY 421 ProThrAlaValAlaAlaValGluThrPheAspArgGlyLysAspPheGlnGlySerLysLeu 440

```

Db      1304 CCCACTGCCAAGCGCTCCGTGGAATGTTGATGGAAGAATTTTCAGAGGACCAACTT 1363
Qy      441 LysValSerLeuAlaArgIlylvsProProMetAsnSerMetArgGlyIylLeuProPro 460
Db      1364 AAAGTCTCCCTTGTCTGGAGAAAGCCCTCCAAATGAATGAGTGGGGGTGTCTGCAACC 1423
Qy      461 ArgGluIylArgGlyMetProProProLeuAlaArgIylIylProGlyIylPro 480
Db      1424 CGTGAGGGCAGAGGCAAGCCACCACTCGTGGAGAGTTCAGAGGCGCAGAGGCTCT 1483
Qy      481 GlyIylProMetGlyIylArgMetGlyIylArgGlyIylAspArgGlyIylPheProProArg 500
Db      1484 GGGGGACCCCATGGGTCCGATGGAGGCGCTGGAGAGATAGAGAGGCTTCCCTCCCAAGA 1543
Qy      501 GlyProArgGlyIylSerArgGlyIylAsnProSerGlyIylIylIylAsnValGlnIylAspAlaGly 520
Db      1544 GAGACCCCGGGGATTCCTCCAGAGAACCTCTGGAGAGAGAAACGTCAGACCGAGCTGGA 1603
Qy      521 AspTrpGlnIylCysProAsnProGlyIylCysGlyIylAsnGlnAsnPheAlaTrpArgTrpGlnIylCys 540
Db      1604 GACTGGCAGTGTCCCATTCGGGTGTGGAAACCAAACTTCGCTGGAGAACAGAGTGC 1663
Qy      541 AsnGlnIylCysAlaIylAspIylProGlyIylPheLeuProProPheProProGlyIyl 560
Db      1664 AACCAAGTAAAGGCCCAAGACCTGAAAGGCTTCTCCCGCCACCTTCCGCGCGGT 1723
Qy      561 GlyAspArgGlyIylArgGlyIylProGlyIylMetArgGlyIylIylArgGlyIylLeuMetAsp 580
Db      1724 GGTATGTGTGCAAGAGTGTGCTGTGGCAATGGGAGAGAAAGTGTGCTCATGAT 1783
Qy      581 ArgGlyIylProGlyIylMetPheArgGlyIylIylArgGlyIylAspArgGlyIylPheArg 600
Db      1784 CGTGTGTGTCCCGTGGAAATTTTCAAGGTGGCCGTGTGGAGACAAAGTGTGCTTCCGT 1843
Qy      601 GlyIylArgGlyIylMetAspArgGlyIylIylPheGlyIylIylArgGlyIylProGlyIyl 620
Db      1844 GGTGGCGGGGCAATGACCGAGGTGGCTTGTGTGAGAAAGACGAGGTGGCGCGGG 1903
Qy      621 ProProGlyIylProLeuMetGlnIylMetGlyIylIylArgGlyIylIylArgGlyIylProGlyIyl 640
Db      1904 CCCCCTGACCTTTGATGAGAAACAGATGGAGAAAGAGAGACGTGAGACCTGGA 1963
Qy      641 LysMetAspIylGlyIylIylAspGlnIylIylArgArgAspArgProIyl 656
Db      1964 AAAATGATTAAGCGCAGACCCGTCAAGAGCGCAGAGATCGCCCTAC 2011

```

RESULT 7  
ACN40903  
ID ACN40903 standard; cDNA, 2390 BP.

```

XX      ACN40903;
XX      18-NOV-2004 (first entry)
XX      Tumour-associated antigenic target (TAT) cDNA DNA269830, SEQ ID NO:5989.
XX      Tumour-associated antigenic target; TAT; human; overexpression; cancer;
XX      tumour; diagnosis; cell proliferative disorder; breast cancer;
XX      colorectal cancer; lung cancer; ovarian cancer; liver cancer;
XX      central nervous system cancer; bladder cancer; pancreatic cancer;
XX      cervical cancer; melanoma; leukaemia; hybridisation probe;
XX      chromosome identification; chromosome mapping; gene mapping;
XX      gene therapy; cytotoxic; gene; ss.
XX      Homo sapiens.
XX      OS
XX      PN
XX      MO2004030615-A2.
XX      15-APR-2004.
XX      29-SEP-2003; 2003WO-US028547.
XX      02-OCT-2002; 2002US-041971P.

```

```

XX      (GETH) GENENTECH INC.
XX      Wu TD, Zhang Z, Zhou Y;
XX      MPI; 2004-347921/32.
XX      P-PSDB; ABM82330.
XX      New tumor-associated antigenic target polypeptides and nucleic acids,
XX      useful in preparing a medicament for treating or detecting a
XX      proliferative disorder, e.g. breast, lung, colorectal, ovarian or
XX      prostate cancer or tumor.
XX      Claim 1; SEQ ID NO 5989; 7273bp; English.
XX      The invention relates to human tumour-associated antigenic target (TAT)
XX      polypeptides, and their related nucleic acids. The TAT polypeptides are
XX      overexpressed in cancer tissues compared to normal tissues, and may thus
XX      serve as effective targets for the diagnosis and treatment of cancer in
XX      mammals. The invention also relates to nucleic acid and polypeptide
XX      sequences at least 80% identical to the TAT nucleic acids and
XX      polypeptides; expression vectors and host cells comprising a TAT nucleic
XX      acid; an antibody specific for a TAT polypeptide; a peptide or organic
XX      molecule which binds to a TAT polypeptide; fusion proteins comprising a
XX      TAT polypeptide; and methods and compositions for the treatment or
XX      diagnosis of cancer in mammals. TAT polypeptides, nucleic acids,
XX      antibodies, antagonists, binding molecules and compositions are useful
XX      for diagnosing or treating a cell proliferative disorder associated with
XX      increased TAT expression, particularly cancers such as breast cancer,
XX      colorectal cancer, lung cancer, ovarian cancer, liver cancer, bladder
XX      cancer, pancreatic cancer, cervical cancer, cancers of the central
XX      nervous system, melanoma and leukaemia. TAT nucleic acids may further be
XX      used as hybridisation probes, in chromosome and gene mapping, in
XX      chromosome identification and in gene therapy. The present sequence
XX      represents a TAT nucleic acid of the invention.
XX      Sequence 2390 BP; 645 A; 589 C; 668 G; 488 T; 0 U; 0 Other;
XX      Alignment Scores:
XX      Pred. No.: 1,7e-116 Length: 2390
XX      Score: 3633.00 Matches: 656
XX      Percent Similarity: 100.00% Conservative: 0
XX      Best Local Similarity: 100.00% Mismatches: 0
XX      Query Match: 100.00% Indels: 0
XX      DB: 13 Gaps: 0
XX      US-10-791-017A-2 (1-656) x ACN40903 (1-2390)
Qy      1 MetAlaSerThrAspIylSerThrIylSerGlnAlaAlaIylGlnIylIylSerAla 20
Db      44 ATGGCGTCCACGGAATTACAGTACTTATAGCAAGCTGACGAGCGGCTACAGTGTCT 103
Qy      21 TyrThrAlaGlnProThrGlnIylIylAlaGlnThrThrGlnAlaIylIylGlnIylSer 40
Db      104 TACACCGCCCGACCCACTCAAGATATGCAAGACCACCCAGGCATATGGCAACAAAGC 163
Qy      41 TyrGlyIylThrIylGlnIylProThrAspValSerIylThrGlnAlaGlnThrThrAlaThr 60
Db      164 TATGAAACCTATAGACAGCCCACTGATGTCAGTATATACCAAGCTCAGACACTGCAACC 223
Qy      61 TyrGlyIylThrAlaIylThrAlaIylThrSerIylGlyIylGlnIylProProThrGlyIylThrThrPro 80
Db      224 TATGGCGAGACCCCTATGCACTTCTTATGCAAGCTCCCACTGGTTATATCACTCCA 283
Qy      81 ThrAlaProGlnAlaIylSerGlnIylProValGlnIylIylIylThrGlyIylAlaIylThr 100
Db      284 ACTGCCCCCGAGGATATACAGCCAGCTGTCCAGAGGGTATATGCACTGTATATATACC 343
Qy      101 ThrThrAlaThrValThrThrThrGlnAlaIylThrAlaIylThrAlaIylThrThrThr 120
Db      344 ACCACTGTATACATCCACCAACAGGCTCTTATGCAAGCTCAGTGTGATATATGGACT 403
Qy      121 GlnProAlaIylProAlaIylThrGlnIylGlnIylProAlaIylThrAlaIylThrThrArgProGln 140

```



CC derived from a variety of human tissues. The invention also provides a  
CC method for producing proteins from these polynucleotide sequences. The  
CC proteins are useful for identifying compounds that modulate their  
CC activity and production, and the cell is also useful for identifying  
CC compounds that modulate expression of the polynucleotide sequences  
CC encoding the secreted proteins. The sequences of the invention are useful  
CC for treating diseases such as hyperproliferative disorders (e.g. cancer),  
CC immune deficiency disorders (e.g. severe combined immunodeficiency  
CC (SCID)), autoimmune disorders (e.g. multiple sclerosis), blood disorders  
CC (e.g. thrombocytopenia), inflammatory disorders (e.g. arthritis) and  
CC infectious disorders (e.g. hepatitis). The polynucleotide sequences of  
CC the invention are also useful in gene therapy. A562214-A562388  
CC represent the cDNA sequences of the invention that encode for novel human  
CC secreted proteins  
XX  
XX  
XX Sequence 2176 BP, 413 A, 617 C, 563 G, 583 T, 0 U, 0 Other;  
50

**SQ Sequence 2176 BP; 413 A; 617 C; 563 G; 583 T; 0 U; 0 Other;**

**Alignment Scores:**

Pred. No.:	1,446,115	Length:	217
Score:	3605.00	Matches:	655
Percent Similarity:	99.85	Conservative:	0
Best Local Similarity:	99.85	Mismatches:	1
Query Match:	99.23	Indels:	1
DB:	6	Gaps:	0

US-10-791-017A-2 (1-656) x AAS62262 (1-2176)

Qy	MetLlaSerThrAspPylSerSerThrTyrSerGlnIleAlaIleGlnGlnGlyTyrSerLeu	20
Db	ATGGGCTCCAGGATTTACAGTACTTATAGCCAGCTGCAGGCGACAGAGGGCTACAGTCT	2101
Qy	21 TyrThrAlaGlnProThrGlnGlyTyrAlaGlnThrThrGlnAlaTyrGlyGlnGlnSer	40
Db	2100 TACACCGCCAGCCCACTCAAGGATATGCAAGACCACTCCAGGCAATATGGGCAACAAAG	2041
Qy	41 TyrGlyThrTyrGlyGlnProThrAspValSerTyrThrGlnAlaGlnThrThrAlaThr	60
Db	2040 TATGGAACCTTATGAGCAGCCCACTGATCTCAGCTTATACCCAGGCTCAACACCACTGGCAAC	1981
Qy	61 TyrGlyGlnThrAlaTyrAlaThrSerTyrGlyGlnProProThrGlyTyrThrThrPro	80
Db	1980 TATGGCGAGCCGCGCTATGCAACTCTTATATGACAGCTCCCACTGGTATATCTACTCA	1922
Qy	81 ThrAlaProGlnAlaTyrSerGlnProValGlnGlyTyrGlyThrGlyAlaTyrAspThr	100
Db	1920 ACTCCCCCCAGGCAATACAGCCAGCTCTCCAGGGGTATGGCACTGGTCTTATGATAC	1861
Qy	101 ThrThrAlaThrValThrThrThrGlnAlaSerTyrAlaAlaGlnSerAlaTyrGlyThr	120
Db	1860 ACCACTGTATAGTCACACCAACCCAGGCTCTCTATGAGCTCACTGCAATGCAATATGGCACT	1801
Qy	121 GlnProAlaTyrProAlaTyrGlyGlnGlnProAlaAlaThrAlaProThrAspProGln	140
Db	1800 CAGCTGCTTATCCAGCTTATGGCGAGCAGCAGCCAGCCCACTGCACCTTACAGACCCGA-	1744
Qy	141 AspGlyAsnLysProThrGlnLuthSerGlnProGlnSerSerThrGlyGlyTyrAsnGln	160
Db	1741 GATGGAACCAAGCCCACTGAGACTAGTCAACTCAATCTAGACAGGGGGTTACAAACAG	1682
Qy	161 ProSerLeuGlyTyrTyrGlyGlnSerAsnTyrSerTyrProGlnValProGlySerTyrPro	180
Db	1681 CCCAGCCTCAGATATGACAGAGTAACTAACAGTTATCCCAAGTATCCGGAGACTAACCC	1622
Qy	181 MetGlnProValThrAlaProProSerTyrProProThrSerTyrSerSerThrGlnPro	200
Db	1621 ATGCAAGCAGTCACTGCACCTTCAATCTTACCTCTTACAGCTATCTCTTACACAGCCG	1566
Qy	201 ThrSerTyrAspGlnSerSerSerTyrSerGlnGlnAsnThrTyrGlyGlnProSerSerTyr	220
Db	1561 ACTAGTTATATGATCAGAGCACTTACTCTCAGCAGAACCTATATGGGCAACCGAGAGCTAT	1502
Qy	221 GlyGlnGlnSerSerTyrGlyGlnGlnSerSerTyrGlyGlnGlnProProThrSerTyr	240

Dd	1501	GGA	CAG	CAG	AGT	AGT	AGT	CA	CA	AA	GA	GC	TAT	TGG	CA	GC	AG	CCT	CC	CA	TAA	TT	AC	1442					
Oy	241	Pro	Pro	Gln	Thr	Thr	Gly	Ser	Thr	Ser	Gln	Ala	Pro	Ser	Gln	Thr	Ser	Gln	Thr	Ser	Ser	Ser	Ser	260					
Dd	1441	CC	ACC	CC	CA	AA	TGG	AT	CT	CA	GC	CA	GC	CA	CT	CA	AA	GT	CA	AA	TA	TAT	GC	CA	AG	CA	GC	AG	1382
Oy	261	Tyr	Gly	Gln	Gln	Ser	Ser	Phe	Arg	Gln	Asp	His	Pro	Ser	Ser	Met	Gly	Val	Thr	Gly	Gln				280				
Dd	1381	TAC	GGG	CA	CA	CA	GAT	TAT	TCC	CA	CA	GA	CA	CA	CC	CA	GTA	AG	CAT	GGT	GTT	TAA	GGG	CAG	1322				
Oy	281	Glu	Ser	Gly	Phe	Ser	Gly	Pro	Gly	Val	Asn	Arg	Ser	Met	Ser	Gly	Pro	Asp	Asn	Arg					300				
Dd	1321	GAG	CT	CA	GA	GA	AT	TTT	CT	CC	GA	CC	CA	GA	GA	AA	CC	GG	CA	TGA	GT	GC	CT	CA	TAA	CC	GG	1262	
Oy	301	Gly	Val	Arg	Gly	Val	Arg	Phe	Asp	Arg	Gly	Val	Met	Ser	Arg	Gly	Val	Arg	Gly	Val					320				
Dd	1261	GG	CAG	GGG	AA	AGG	GGG	AT	TGA	TTC	GT	GA	GC	CA	TGA	GC	CA	AG	AT	TGG	CCG	GGG	AG	GGA	1202				
Oy	321	Arg	Gly	Val	Met	Gly	Ser	Ala	Gly	Val	Arg	Gly	Phe	Asn	Val	Pro	Gly	Val	Pro	Met					340				
Dd	1201	CG	CGT	GT	GA	AT	TGG	CA	CG	CT	GA	GC	AG	CA	AG	TGG	CT	CA	ATA	AG	CT	TGG	TGA	CC	AT	1142			
Oy	341	Asp	Gly	Val	Pro	Asp	Leu	Asp	Leu	Gly	Pro	Val	Asp	Pro	Asp	Gly	Asp	Ser	Asp	Asn					360				
Dd	1141	GAT	AA	AG	CA	CC	AG	AT	CT	TGA	AT	CT	TAA	GC	CA	CT	GT	AA	T	CA	AG	T	CA	GA	CA	CT	CA	CA	1082
Oy	361	Ser	Ala	Thr	Val	Gln	Gly	Leu	Asn	Asp	Ser	Val	Thr	Leu	Asp	Ala	Leu	Asp	Phe						380				
Dd	1081	AGT	CA	AT	TT	TAT	AT	CA	CA	AG	AT	TAA	TA	GA	CA	AG	TGT	CA	T	AG	TGT	AT	CT	GT	CA	CA	CT	1022	
Oy	381	Phe	Val	Gln	Cys	Gly	Val	Val	Leu	Met	Asn	Val	Arg	Thr	Gly	Gln	Pro	Met	Leu	His	Ala				400				
Dd	1021	TTT	AA	GA	CA	GT	TGG	GGT	GT	TTA	AG	AT	TA	CA	AG	AA	CT	TGG	CA	CC	CA	CC	CA	AT	CA	AT	962		
Oy	401	Tyr	Leu	Asp	Val	Gly	Thr	Gly	Val	Pro	Gly	Val	Asp	Ala	Thr	Val	Ser	Tyr	Gly	Asp	Pro				420				
Dd	961	TAC	CT	GA	CA	AA	GA	AA	CA	GA	AA	AG	CC	CA	AA	AG	GC	AT	CC	CA	AG	CT	CT	TAA	GA	AA	CCA	902	
Oy	421	Pro	Thr	Ala	Val	Val	Ala	Val	Gly	Thr	Phe	Asp	Gly	Val	Asp	Phe	Gln	Gly	Ser	Val	Leu				440				
Dd	901	CCC	ACT	CC	CA	AA	GG	CT	CG	CT	CG	TA	AG	TG	TAT	GT	GA	TGG	CA	AA	AT	TTT	CA	AG	GA	CA	AA	CTT	842
Oy	441	Lys	Val	Ser	Leu	Ala	Gly	Ser	Val	Pro	Pro	Met	Asn	Ser	Met	Arg	Gly	Val	Leu	Pro	Pro				460				
Dd	841	AAA	AT	CT	CC	CT	CT	CG	CA	AA	GA	AG	CT	CA	TAA	CA	AG	TA	AG	CG	GG	TGG	CT	CA	CA	CC	782		
Oy	461	Arg	Gly	Gln	Val	Arg	Gly	Met	Pro	Pro	Pro	Leu	Arg	Gly	Gly	Pro	Gly	Gly	Pro						480				
Dd	781																												





```

Db 1074 GATGAAGACCAAGTCTTGATCTAGGCTTCCATATATCCGATGAAGCTCGACAC 1133
Qy 361 SerAlaIeTyValGInGlyLeuAsnAapSerValThreUasAapLeuAlaAapPhe 380
Db 1134 AGTGCATTTATGTGCAAGATTTAAATGACATATGACTCTGATGATCTGGCAGACTTC 1193
Qy 381 PheUasGInCyseGlyValAlaValMetAasPheAaGrhGlyGInPromeCilLeAlle 400
Db 1194 TTTAAGAGATGTGGGTGTGTCAAGATGAACAAGAACTGACAACTGATCATATTC 1253
Qy 401 TyLeuAapPheGInThrGlyLeuPProLyseGlyAaPAlaThreValSerTyGlyUaAapPro 420
Db 1254 TACTTGATTAAGAGACAGAGAAAGCTTAAGGGGAGCCCAAGTCTCTATGAAGATCCA 1313
Qy 421 ProThraAlaValAlaAlaValGluTrpPheAapGlyLeuAapPheGInGlySerTyLeu 440
Db 1314 CCAACTGCAAAAGCTGCGGTGGAATGCTTGATGGAAGATTTTCAAGGAAGCAAACTT 1373
Qy 441 TyValSerLeuAlaArgTyLeuPProPromeAasPheAaGrhGlyLeuPPro 460
Db 1374 AAAGTGTCTTGTGCGCAAAAGAGCTTCGAATGAACAGCATGCGGGAGGATCCACT 1433
Qy 461 ArgGInGlyArgGlyMetProPProPProLeuArgGlyGlyProGlyGlyPro 480
Db 1434 CGTGAAGGCGAGGGGTATGCCACCACTCTGTGAGGTCTGTGGCCAGAGAGGCTT 1493
Qy 481 GlyGlyPromeGlyArgMetGlyGlyArgGlyGlyAaPArgGlyGlyPhePProAaGr 500
Db 1494 GGAAGAGACCACTGCTCCATGCGAGGCGGTGGAAGAGAGAGAGGAGGCTTCCCTCAAGA 1553
Qy 501 GlyPProAaGrGlySerArgGlyAaPProSerGlyGlyGlyAaPValGInAlaArgAlaGly 520
Db 1554 GGGGCGGAGGCTCAAGAGAAAGCCCTCTGAGAGAGAAATGTCACAGCAGGAGCTGGA 1613
Qy 521 AaPTrpGInCyPProAaPProGlyGlyArgGlyAaGInAaPheAlaTrpArgThruCyS 540
Db 1614 GACTGCGAGTCTCCCATTCGGGCTGTGGAACCAAGAACTTCTGTGGAAGAGCAAGATTC 1673
Qy 541 AaGInCyValAaPProLysePProGInGlyPheLeuPProPProPProPProGly 560
Db 1674 AACCAAGTAAAGGCGCCCTAAGCCGAGGGCTTCTCCCGCACCTTCCACTCCGGGT 1733
Qy 561 GlyAaPArgGlyAaGrGlyGlyProGlyGlyMetAaGrGlyGlyAaGrGlyGlyLeuMetAaP 580
Db 1734 GGTATGCTGAACAGAGGTGCGCTGTGTGCAATGAGAGAGAGAGAGAGAGAGAGAGAG 1793
Qy 581 ArgGlyGlyProGlyGlyMetPheAaGrGlyGlyAaGrGlyGlyAaPArgGlyPheAaGr 600
Db 1794 CTTGTGTCTCTGAGAGAAATTTCAAGAGTGTGAGAGAGTGTGAGAGAGAGAGAGAGAG 1853
Qy 601 GlyGlyArgGlyMetAaPArgGlyGlyPheGlyGlyGlyAaGrGlyGlyProGlyGly 620
Db 1854 GGTGCGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1913
Qy 621 ProPProGlyPProLyseMetGlyGlyMetGlyGlyAaGrGlyGlyAaGrGlyGlyProGly 640
Db 1914 CTTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1973
Qy 641 TyMetAaPArgGlyGlyGlyMetAaGrGlyGlyGlyGlyGlyAaGrGlyGlyProGly 656
Db 1974 AAAATGATTAAGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2021

```

```

Km Human; chromosome mapping; gene mapping; gene therapy; forensic;
Kw food supplement; medical imaging; diagnostic; genetic disorder; ss.
Xx Homo sapiens.
Xx MO200175067-A2.
Xx 11-OCT-2001.
Xx 30-MAR-2001; 2001MO-US008631.
Xx 31-MAR-2000; 2000US-00540217.
Xx 23-AUG-2000; 2000US-00649167.
Xx (HYSE-) HYSEQ INC.
Xx Dmanac RT, Liu C, Tang YT;
Xx WPI; 2001-639362/73.
Xx P-PDB; ABG06460.
Xx New isolated polynucleotide and encoded polypeptides, useful in
Xx diagnostics, forensics, gene mapping, identification of mutations
Xx responsible for genetic disorders or other traits and to assess
Xx biodiversity.
Xx Claim 1; SEQ ID NO 6451; 103bp; English.
Xx The invention relates to isolated polynucleotide (I) and polypeptide (II)
Xx sequences. (I) is useful as hybridisation probes, polymerase chain
Xx reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
Xx and in recombinant production of (II). The polynucleotides are also used
Xx in diagnostics as expressed sequence tags for identifying expressed
Xx genes. (I) is useful in gene therapy techniques to restore normal
Xx activity of (II) or to treat disease states involving (II). (II) is
Xx useful for generating antibodies against it, detecting or quantitating a
Xx polypeptide in tissue, as molecular weight markers and as a food
Xx supplement. (II) and its binding partners are useful in medical imaging
Xx of sites expressing (II). (I) and (II) are useful for treating disorders
Xx involving aberrant protein expression or biological activity. The
Xx polypeptide and polynucleotide sequences have applications in
Xx diagnostics, forensics, gene mapping, identification of mutations
Xx responsible for genetic disorders or other traits to assess biodiversity
Xx and to produce other types of data and products dependent on DNA and
Xx amino acid sequences. Aa664197-AA94564 represent novel human diagnostic
Xx coding sequences of the invention. Note: The sequence data for this
Xx patent did not appear in the printed specification, but was obtained in
Xx electronic format directly from WIPO at
Xx ftp.wipo.int/pub/published_pct_sequences
Xx
Xx SO Sequence 2388 BP; 628 A; 599 C; 668 G; 493 T; 0 U; 0 Other;
Xx
Xx Alignment Scores:
Xx Pred. No.: 5.11e-113 Length: 2388
Xx Score: 3531.50 Matches: 651
Xx Percent Similarity: 98.64% Conservative: 2
Xx Best Local Similarity: 98.34% Mismatches: 3
Xx Query Match: 97.21% Indels: 6
Xx DB: 5 Gaps: 1
Xx
Xx US-10-791-017a-2 (1-656) x AAS70647 (1-2388)
Xx
Xx Qy 1 MetAlaSerThrAaPProLyseTrpTySerGInAlaAlaGInGInGlyTyTrSerAla 20
Xx Db 53 ATGGCGTCAAGGATTAAGTACTTAAAGCAAGCTGACAGAGAGAGAGAGAGAGAGAGAG 112
Xx Qy 21 -TyThraAlaGInProThrGInGlyTyTrAlaGInThrThr---GInAlaTyGInGInG 39
Xx Db 113 TTACACCGCCAGCCCACTCAAGATATGACAGACCAAGAGAGAGAGAGAGAGAGAGAGAG 172
Xx Qy 39 nSerTyGlyThrTyGlyGInProThrAaPValSerTyThrGInAlaGInThrThra 59
Xx Db 173 AAGCTATGAAGAACTATGACAGCCCACTGATGTCAAGTATACCAAGGCTGACAGACTGC 232

```

QY 59 eThrTyrglyqlnThrAlaTyraIaThrSerTyrglyqlnProProthrglyTyrrTh  
 Db 233 AACCTAAGGGGAGACCGCTATGCAACTTCTTATGCAAGCCTCCACTGGTATACTAC 292  
 QY 79 rProThraIaProGlnaIaTySer-GlnProValGln-GlyTyrglyThglAlaTy 98  
 Db 293 TCCAACTGCCCCCAAGGATACAGCCAGCCTGTCAGGGGATAGGCACTGGTCTAT 352  
 QY 99 AspThrThraIaThrValThr-ThrThrglnaIaSerTyraIaIaGlnSerAlaTy 118  
 Db 353 GATACCACTGCTTACAGTCAACCAACCAAGCCCTCCATGCAAGTCACTGGATA 412  
 QY 118 rglTyThglnProAlaTyProAlaTyrglyqlnGlnProAlaIaIaThraIaProthar 138  
 Db 413 TGGCACTCAAGCTCTTATCAAGCTATGAGCAGCAGCAGCAAGCCTGCTTACAG 472  
 QY 138 gProGlnAsp-GlyAsnIaSerProthrglyThrglnProGlnSerSerThrglyqln 158  
 Db 473 AACCGAAGATTGAAACCAAGCCCACTGAGACTAGTCAACTCAATCTAGCAAGAGGCTT 532  
 QY 158 yraAnglnProSerleuGlyTyrglyqlnSerAenTySerTyProGlnAlaProGlys 178  
 Db 533 ACAACCAAGCCCAAGCTTGAATGAGACAGAGTAAGTCAAGTATCCCAAGTACTGGGA 592  
 QY 178 eTyProMetGlnProValThraIaProProSerTyProProthSerTySerSerT 198  
 Db 593 GCTAACCCATGCAAGCTCACTGCTCCTCCTCAAGCTTCAAGCTTCAAGCTTCAAG 652  
 QY 198 hrglnProthSerTyAspGlnSerSerTySerGlnIaThraIaThrglyqlnPro 218  
 Db 653 CACAGCGAAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 712  
 QY 218 eSerTyrglyqlnGlnSerSerTyrglyqlnGlnSerSerTyrglyqlnGlnPro 238  
 Db 713 GCAAGCTATGAGCAGCAGATAGTATGATGATGATGATGATGATGATGATGATGATG 772  
 QY 238 hSerTyProProGlnThrglyqlnSerTySerGlnIaProSerGlnTySerGln 258  
 Db 773 CTAGTATCCCAAGCTTGAATGATGATGATGATGATGATGATGATGATGATGATG 832  
 QY 258 eSerSerTyrglyqlnGlnSerSerPheArgGlnaIaProSerSerThrglyqln 278  
 Db 833 GCAAGCTATGAGCAGCAGATAGTATGATGATGATGATGATGATGATGATGATGATG 892  
 QY 278 yrglyqlnGlnSerGlyPheSerGlyProGlyqlnAenArgSerMetSerGlyPro 298  
 Db 893 ATGGGCAAGAGTCTGAGGATTTTCCGAGCAGAGAGAGAGAGAGAGAGAGAGAGAG 952  
 QY 298 sPaAsnArgGlyArgGlyArgGlyqlnIaPheAspArgGlyqlnMetSerArgGly 318  
 Db 953 ATAAACCGGGGAG 1012  
 QY 318 yglIyglIyArgGlyqlnMetGlySerIaGlyqlnArgGlyqlnPheAsnIaSer 338  
 Db 1013 GAGGAG 1072  
 QY 338 lProMetAspGlnGlyProAspIaPheArgGlyqlnProValaIaProAspGln 358  
 Db 1073 GACCGAG 1132  
 QY 358 ePaAsnArgGlyqlnIaIaTyValGlnIaIaAsnAspSerValThraIaPheAsp 378  
 Db 1133 CTGACCAACAGTGCATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1192  
 QY 378 lAspPhePheIa 398  
 Db 1193 CAGACTCTTTAAGCAGAGT 1252  
 QY 398 lEhIa 418  
 Db 1253 TCCAGACTCACTCTGAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1312

QY 418 lAspProProthraIa 438  
 Db 1313 AAGACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC 1372  
 QY 438 eTyIa 458  
 Db 1373 GCAACTTAAAGTCT 1432  
 QY 458 eProProArgGlnIa 478  
 Db 1433 TGGCAACCCGAG 1492  
 QY 478 yglIyProGlyqlnProMetGlyArgMetGlyqlnArgGlyqlnIaPheArg 498  
 Db 1493 GAGGTCTGGGGGAGCCATGAGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1552  
 QY 498 rProArgGlyProArgGlySerArgGlyIaAsnProSerGlyqlnIaIaIaIaIaIa 518  
 Db 1553 CTCCAG 1612  
 QY 518 rgaIa 538  
 Db 1613 GAGCTGAG 1672  
 QY 538 hrgIuCyAsnGlnCyAsnIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 558  
 Db 1673 CAGAGTCAACCAAGTGTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1732  
 QY 558 rProArgGlyIa 578  
 Db 1733 CCGGAG 1792  
 QY 578 eMetAspArgGlyqlnProGlyqlnMetPheArgGlyqlnArgGlyqlnIaIaIaIa 598  
 Db 1793 TCATGATCGAG 1852  
 QY 598 lPheArgGlyqlnArgGlyqlnMetAspArgGlyqlnIaIaIaIaIaIaIaIaIaIa 618  
 Db 1853 GCTTCCGT 1912  
 QY 618 rglIyglIyProProGlyqlnProMetGlnIaIaIaIaIaIaIaIaIaIaIaIaIaIa 638  
 Db 1913 CTGG 1972  
 QY 638 lProGlyqlnMetAspIa 656  
 Db 1973 GACCTGGAAG 2028  
 RESULT 11  
 AAS62623/c  
 ID AAS62623 standard; cDNA; 2273 BP.  
 XX  
 AC AAS62623;  
 XX  
 DT 14-FEB-2002 (first entry)  
 DE  
 XX cDNA sequence #410 encoding novel human secreted protein.  
 XX  
 KW Human secreted protein; hyperproliferative disorder; autoimmune disorder;  
 KW immune deficiency disorder; blood disorder; inflammatory disorder;  
 KW infectious disorder; gene therapy; antimicrobial; hepatotropic;  
 KW immunosuppressive; antineumatic; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN NO200177291-A2.  
 XX  
 PD 18-OCT-2001.  
 XX  
 PF 29-MAR-2001; 2001MO-US010485.  
 XX  
 PR 06-APR-2000; 2000US-0195604P.  
 XX

PA (GENY ) GENETICS INST INC.

XX Wong CG, Clark HF, Reichel K, Agostino MJ, Howes SH, Resnick RJ;  
PI Gulukota K, Graham JR;  
XX WPI, 2002-010900/01.

XX New polynucleotides encoding secreted proteins useful for treating e.g.  
PT asthma, HIV and Crohn's disease.

PS Claim 1, Page 291, 391pp; English.

XX The present invention relates to the isolation of novel cDNA sequences  
CC which encode human secreted proteins. The cDNA sequences have been  
CC derived from a variety of human tissues. The invention also provides a  
CC method for producing proteins from these polynucleotide sequences. The  
CC proteins are useful for identifying compounds that modulate their  
CC activity and production, and the cell is also useful for identifying  
CC compounds that modulate expression of the polynucleotide sequences  
CC encoding the secreted proteins. The sequences of the invention are useful  
CC for treating diseases such as hyperproliferative disorders (e.g. cancer),  
CC immune deficiency disorders (e.g. severe combined immunodeficiency  
CC (SCID)), autoimmune disorders (e.g. multiple sclerosis), blood disorders  
CC (e.g. thrombocytopenia), inflammatory disorders (e.g. arthritis) and  
CC infectious disorders (e.g. hepatitis). The polynucleotide sequences of  
CC the invention are also useful in gene therapy. AAS62214-AAS62838  
CC represent the cDNA sequences of the invention that encode for novel human  
CC secreted proteins

XX Sequence 2273 BP; 477 A; 636 C; 562 G; 598 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 2,26e-111 Length: 2273  
Score: 3483.00 Matches: 636  
Percent Similarity: 97.10% Conservative: 0  
Best Local Similarity: 97.10% Mismatches: 1  
Query Match: 95.87% Indels: 18  
DB: 6 Gaps: 2

US-10-791-017a-2 (1-656) x AAS62623 (1-2273)

QY 2 AAlaSerThrAspYrSerThrYrSerGlnAlaAlaGlnGlnGlyYrSerAlaYr 21  
Db 2263 GCGTCCACGATTACAGTACTATAGCCAGCTCAGCGCAGCGGCTTACAGCTTAC 2264  
QY 22 ThrAlaGlnProThrGlnGlyYrAlaGlnThrThrGlnAlaYrGlnGlnSerYr 41  
Db 2203 ACCGCCAGCCACTCAAGATATGACAGACCAACCCAGCATATGGGCAACAAAGCTAT 2144  
QY 42 GlyThrYrGlnGlnProThrAspValSerYrThrGlnAlaGlnThrThrAlaThrYr 61  
Db 2143 GGAACCTATGAGACGCCACTGATGTGCTATACCCAGGCTCAAGACCACTGCACTTAT 2084  
QY 62 GlyGlnThrAlaYrAlaThrSerYrGlnGlnProThrGlnYrThrThrProThr 81  
Db 2083 GGGGAGACCGGCTATGCAACTTCTTATGACAGCTCCCACTGTTATACCTCACT 2024  
QY 82 AlaProGlnAlaYrSerGlnProValGlnGlnGlyYrGlnYrAlaYrAspThrThr 101  
Db 2023 GCCCCCCAGGATACAGCCAGCTGTCCAGGGGTATGACCTGGGTATGATACCAACC 1964  
QY 102 ThrAlaThrValThrThrThrGlnAlaSerYrAlaAlaGlnSerAlaYrGlnYrThnGln 121  
Db 1963 ACTGCTATAGTACCAACCAACCGAGCTCTTATGAGCTCAAGTCTGCAATATGGCACTCAG 1904  
QY 122 ProAlaYrProAlaYrGlnGlnProAlaAlaThrAlaProThrArgProGlnAsp 141  
Db 1903 CCTGCTTATCAAGCTATGAGGAGACAGCAGCAGCACTGCACTTACAAAGCCAGAGAT 1844  
QY 142 GlyAsnYrProThrGlnThrSerGlnProGlnSerSerThrGlnYrYrAsnGlnPro 161  
Db 1843 GGAACCAAGCCCACTGAGACTATGATCACTCATCTATGACAGGGGGTATACAAACAGCCC 1784

QY 162 SerLeuGlyYrGlnSerAsnYrSerYrProGlnValProGlySerYrPromet 181  
Db 1783 AGCTTAGATATGAGACAGATATACAGTATATCCCAAGTACTGGAGCTACCCCATG 1124  
QY 182 GlnProValThrAlaProProSerYrProProThrSerYrSerSerThrGlnProThr 201  
Db 1723 CAGCCAGTCACTGCACTCCATCTTACCTCCACAGCATATTCCTTACAGCCGACT 1664  
QY 202 SerYrAspGlnSerSerYrSerGlnGlnAsnThrYrGlnGlnProSerSerYrGly 221  
Db 1663 AGTTATATGACAGAGCTTATCTTCCAGACACCTATGAGGACCAACCACTATGGA 1604  
QY 222 GlnGlnSerSerYrGlnGlnSerSerYrGlnGlnProProThrSerYrPro 241  
Db 1603 CAGCAGATAGTATGATGATCAACAAACAGCTATGGGACAGAGCTCCCACTATGCCA 1544  
QY 242 ProGlnThrGlySerYrSerGlnAlaProSerGlnYrSerGlnGlnSerSerYr 261  
Db 1543 CCCCAACTGATCCATACAGCCCAAGCTCCCAAGTCAATATAGCCACAGAGCAGACTAC 1484  
QY 262 GlyGlnGlnSerSerPheArgGlnAspHisProSerSerMetGlyValYrGlnGln 281  
Db 1483 GGGCAGCAGAGTTCATTCGACAGACCAACCCAGTAGAGGTGTTATGAGGACAGAG 1424  
QY 282 SerGlyYrPheSerGlyProGlyGlnAsnArgSerMetSerGlyProAspAsnArgGly 301  
Db 1423 TCTGAGAGATTTTCCGACCAAGAGAGAAACCGAGCATGATGAGCTGATACCGGGGC 1364  
QY 302 ArgGlyArgGlyYrPheAspArgGlyGlyMetSerArgGlyYrGlnGlnYrGln 321  
Db 1363 AGCGAAGAGGGGAGATTGATCTGTGAGGCTATGACAGAGGTGGCGGAGAGAGAGCCG 1304  
QY 322 GlyGlyMetGlySerAlaGlyGlnArgGlyYrGlyPheAsnYrProGlyGlyPrometAsp 341  
Db 1303 GGTGAATGGGC--GCTGAGAGCCAGAGGTGGCTTCAATAGCTGTGAGACCAATGAT 1247  
QY 342 GlnGlyProAspLeuAspLeuGlyProProValAspProAspGlyAspAspAsnSer 361  
Db 1246 GAGGACCAAGATCTTATCTAGGCCCACTGTATGATCCAGTGAAGCTGTCAACAGT 1187  
QY 362 AlaIleYrValGlnGlyLeuAsnAspSerValThrLeuAspAspLeuAlaAspPhe 381  
Db 1186 GCAATTTATGTACAGAGATTAATTAATGACAGTGTACTCTGATATCTGGCAGACTCTTT 1127  
QY 382 LysGlnCyArgYrValYrAlaYrMetAsnYrArgThrGlyGlnPrometCileuIleYr 401  
Db 1126 AAGCAGGTGGGTGTGTAAGATGAACAGAGAACTGGGACCACTGATCCACATCTAC 1067  
QY 402 LeuAspLysGlnThrGlyLysProLysGlyAspAlaThrValSerYrGlyLysProPro 421  
Db 1066 CTGAGCAGAGAAACAGAGAAAGCCCAAGGAGTCCACAGTGTCTTATGAAAGCCCAACC 1007  
QY 422 ThrAlaYrAlaAlaValGlnTTPheAspGlyLysAspPheGlnGlySerYrLysLeuYs 441  
Db 1006 ACTGCCAAGGCTCCCTGAGATGTTGTTGGGAAAGATTTTCAAGGAGCAAACTTAAA 947  
QY 442 ValSerLeuAlaArgYrLysProPrometAsnSerMetArgGlyYrLeuProProArg 461  
Db 946 GTCTCCCTGTCTGAGAAAGAGCTTCAATGAACAGTATGGGGGTGTGCAACCCCGT 887  
QY 462 GlyLysArgGlyMetProProLysArgGlyYrProGlyYrGlyProGlyYrProGly 481  
Db 886 GAGGAGAGAGATGCAACCACTCCGTGAGGTTCAGAGGCCCCAGAGAGGTCTGGG 827  
QY 482 GlyProMetGlyArgMetGlyYrArgGlyYrGlyAspArgGlyYrPheProProArgGly 501  
Db 826 GAGACCATGAGGTGCAATGGAGGCGGTGAGAGATAGAGAGGCTTCCCTCAAGAGAGA 767  
QY 502 ProArgGlySerArgGlyAsnProSerGlyYrGlyAsnValGlnHisArgAlaGlyAsp 521  
Db 766 CCCCGGGGTCCGAGAGGAAACCTCTGAGAGGAGAAACCTTCAAGCAGGAGGTGGAGAC 707  
QY 522 TrpGlnCyProAsnProGlyCyGlyAsnGlnAsnPheAlaTrpArgThrGlnCyAsn 541

```

Db      706 TGGCAGTGTCCCATCCGGGTTGTGAAACAGAACTTCGCTGAGAGACAGACTGCAAC 647
Qy      542 GlnCylybAlaProlybProGluGlyPheLeuProProPheProProGlyGly 561
Db      646 CAGGT-----GGT 638
Qy      562 AspArgGlyYArgGlyYProGlyYMetYArgGlyYArgGlyYLeuMetAspArg 581
Db      637 GATGTGGCAGAGGTGGCCCTGGTGGCATGGGGAGAGAGAGGTGGCTTCATGATCGT 578
Qy      582 GlyGlyYProGlyYMetPheArgGlyYArgGlyYArgGlyYAspArgGlyYPheArgGly 601
Db      577 GGTGTCTCCGGTGAATGTTCAAGAGTGGCCGTGTGAGAGCAGAGTGGCTTCGTGT 518
Qy      602 GlyYArgGlyYMetAspArgGlyYArgGlyYArgGlyYArgGlyYProGlyYPro 621
Db      517 GGGCGGGGCAATGGACCGAGGTGGCTTGTGTGAGAGAGACGAGGTGGCCCTGGGGGGCCC 458
Qy      622 ProGlyProLeuMetGluGlnMetGlyYArgGlyYArgGlyYArgGlyYProGlyYLeu 641
Db      457 CTGGACCTTTGATGAGAACAGATGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 398
Qy      642 MetAspLyseGlyGluHisArgGlnGluArgArgAspArgProTyr 656
Db      397 ATGGATAAAGCGAGCAGCCGTCAAGAGCGCAGAGATCGCCCTTAC 353

RESULT 12
AAQ50646
ID      AAQ50646 standard; DNA; 3309 BP.
XX
AC      AAQ50646;
XX
DT      25-MAR-2003 (revised)
DT      26-MAY-1994 (first entry)
XX
DE      Human Ews gene (genomic DNA).
XX
KW      chromosomal translocation; chimeric; chimeric; Ewing sarcoma; Ews gene;
KW      malignant melanoma; hum-fli-1;
KW      primitive peripheral neuroectodermal tumour; human chromosome 11;
KW      human chromosome 22; ss.
XX
OS      Homo sapiens.
XX
FH      Key
FH      exon
FT      1. .37
FT      /+tag= a
FT      /number= 1
FT      /codon_start= 25. .27
FT      38. .96
FT      /+tag= b
FT      /number= 1
FT      97. .133
FT      /+tag= c
FT      /number= 2
FT      134. .193
FT      /+tag= d
FT      /number= 2
FT      194. .245
FT      /+tag= e
FT      /number= 3
FT      246. .305
FT      /+tag= f
FT      /number= 3
FT      306. .429
FT      /+tag= g
FT      /number= 4
FT      430. .489
FT      /+tag= h
FT      /number= 4
FT      490. .676
FT      /+tag= i
FT      /number= 1

```

```

FT      /number= 5
FT      677. .736
FT      /+tag= j
FT      /number= 5
FT      737. .904
FT      /+tag= k
FT      /number= 6
FT      905. .964
FT      /+tag= l
FT      /number= 6
FT      965. .1176
FT      /+tag= m
FT      /number= 7
FT      1177. .1236
FT      /+tag= n
FT      /number= 7
FT      1237. .1417
FT      /+tag= o
FT      /number= 8
FT      1418. .1477
FT      /+tag= p
FT      /number= 8
FT      1478. .1515
FT      /+tag= q
FT      /number= 9
FT      1516. .1575
FT      /+tag= r
FT      /number= 9
FT      1576. .1608
FT      /+tag= s
FT      /number= 10
FT      1609. .1668
FT      /+tag= t
FT      /number= 10
FT      1669. .1787
FT      /+tag= u
FT      /number= 11
FT      1788. .1847
FT      /+tag= v
FT      /number= 11
FT      1848. .1977
FT      /+tag= w
FT      /number= 12
FT      1978. .2037
FT      /+tag= x
FT      /number= 12
FT      2038. .2160
FT      /+tag= y
FT      /number= 13
FT      2161. .2220
FT      /+tag= z
FT      /number= 13
FT      2221. .2383
FT      /+tag= aa
FT      /number= 14
FT      2384. .2443
FT      /+tag= ab
FT      /number= 14
FT      2444. .2541
FT      /+tag= ac
FT      /number= 15
FT      2542. .2601
FT      /+tag= ad
FT      /number= 15
FT      2602. .2854
FT      /+tag= ae
FT      /number= 16
FT      2855. .2914
FT      /+tag= af
FT      /number= 16
FT      2915. .3309
FT      /+tag= ag
FT      /number= 17

```





```

Qy 338 -----GlyPrometAapGlyPro 344
Db 1536 ATAGATATGATATTTATATGATCTTCTGTTGGAGGAGCCATGATGAGAACCA 1595
Qy 345 AapLeuAapLeu----- 348
Db 1596 GATCTGATCTAGAGTAAGTTGAATTCCTAGTTGTCCTTCATATAATTCCTCTGTTG 1655
Qy 349 -----GlyProProvalAapProAapGlyAapSerAapAapSerAapAapLeu 364
Db 1656 TTCTCTCTGAAAGCCCACTGATGATCCAGTATGACCTGCAACAGGCAATTTAT 1715
Qy 365 ValGlnGlyLeuAapSerValThreAapAapLeuAapAapPheLeuGlnCys 384
Db 1716 GTTCAAGGATTAATTAATGACATGATGATCTGAGCACTTCTTAAAGCAATGT 1775
Qy 385 GlyValValLeu----- 388
Db 1776 GGGGTTGTTAAGTCAGTAAAGCATAAACAGTCATCTGGCTCATGCTTAATCTATGCTA 1835
Qy 389 -----MetAapLeuAapThreGlyGlnPheMetLeuLeuLeuLeuLeuLeu 404
Db 1836 TTCTTTCTAGATGAACAAGAGAACCTGGCAACCATGATCCATCTACCTGAGACAG 1895
Qy 405 GluThrGlyLeuProLeuGlyAapAalaThreValSerGlyAapProProThraAly 424
Db 1896 GAAACAGGAAAGCCCAAGGAGATGCAAGTGTCTATGAAGACCAACCATGAGCAAG 1955
Qy 425 AlaAlaValGlyTrpPheAap----- 431
Db 1956 GCTGCGCTGAAATGCTTGAATGAGATGACTCACTGGCATTTCTAATCTAGTAATG 2015
Qy 432 -----GlyLeuAapPheGlnGlySerLeuLeuLeuLeuLeuLeu 444
Db 2016 ATGTTCTGTTGTTCTGTTCCAGGAAAGATTTCCAGGAGCAAACTTAAAGTCTCCCT 2075
Qy 445 AlaArgLeuLeuProPheMetAapSerMetArgGlyLeuProProAapGlyAapArg 464
Db 2076 GCTTCGAAAGAGCTCCATGAATGACATGATGGGGGTGCTGCAACCCCGAGGAGCA 2135
Qy 465 GlyMetProProProLeuAapGly----- 472
Db 2136 GGCATGCAACCACTCCGAGGAGTACTTTACTGAGCTCCATGCTTGAATTTG 2195
Qy 473 -----GlyProGlyLeuProGlyProGlyProGlyProMet 484
Db 2196 CTGTTCTGTTGTTCTGTTGTTGATGATGAGGAGCCAGAGGCTCTGGGAGCCCATG 2255
Qy 485 GlyArgMetGlyGlyAapArgGlyGlyAapArgGlyGlyPheProAapArgGlyProAap 504
Db 2256 GGTTCGATGGAGGAGCCGTTGAGATGAGAGGCTTCCCTCCAGAGAGACCCCGGGGT 2315
Qy 505 SerArgGlyAapProSerGlyGlyAapValGlnIleArgAlaGlyAapTrpGlnCys 524
Db 2316 TCCGAGGAGAACCCCTCTGAGAGAGAAAGTCCAGACCGAGTGAAGTGGCACTGT 2375
Qy 525 ProAapPro----- 527
Db 2376 CCCAATCCGATGATCTGTCGGAATAATGATACCTGATTTCTGCTGATGATTAAT 2435
Qy 528 -----GlyCysGlyAapGlnAapPheAalaThraArgThrGlyCysAapGlnCys 544
Db 2436 GATGCAAGGGGTTGTGAAACAGAACTTCCCTGAGAAACAGAGTCAACAGTGTAAAG 2495
Qy 545 AlaProLeuProGlyGlyPheLeuProProPheProProPro----- 559
Db 2496 GCCCAAGAGCTGAAGCTTCTCTCCGCAACCTTCTCCGCGGGGTAGTGAAGTTT 2555
Qy 560 -----GlyGlyAapAapGly 564
Db 2556 CATGAGTGTCCCTCAACCTCTTCCGCTGATTTCTCACTTAAAGTGTGATCGGAGC 2615
Qy 565 ArgGlyGlyProGlyGlyMetArgGlyGlyArgGlyGlyLeuMetAapArgGlyGlyPro 584

```

```

Db 2616 AGAGTGGCCCTGTGTGCAATGCGGAGAGAGAGTGGCTCATGATCGTGTGCTCC 2675
Qy 585 GlyCysMetPheArgGlyGlyAapArgGlyGlyAapArgGlyGlyPheArgGlyGlyAap 604
Db 2676 GGTGAATGTTTCAAGAGTGGCCCGTGTGAGAGACAGAGTGGCTTCCGTGGTGGCGGGGC 2735
Qy 605 MetAapArgGlyGlyPheGlyGlyGlyAapArgGlyGlyProGlyGlyProProGlyPro 624
Db 2736 ATGACCGAAGTGGCTTGTGTGAGAGAGAGAGTGGCCCTGGGGGCCCCCTGAGACT 2795
Qy 625 LeuMetGlnGlnMetGlyGlyAapArgGlyGlyAapArgGlyGlyProGlyGlyMetAap 644
Db 2796 TTGATGAAACAGATGGAGAGAGAGAGAGAGACGTGAGGACCTGGAATAATGATTAAG 2855
Qy 644 ----- 644
Db 2856 TAAGTCTGGTGAAGAACAGCTGTGGGCTCTAACCGAAGGGCCCTTTACTTGCAGA 2915
Qy 645 GlyGlnHleArgGlnGlnArgAapAapArgProTyr 656
Db 2916 GCGAGCACCGTCAAGAGCCGAGAGATCGGCCCTTAC 2951

RESULT 13
ADA53506
ID ADA53506 standard, cDNA; 1988 BP.
XX
AC ADA53506;
XX
DT 20-NOV-2003 (first entry)
XX
DE Human coding sequence, SEQ ID 1074.
XX
XX Cytostatic; Anti-inflammatory; Osteopathic; Neuroprotective; Nootropic;
KM Gene Therapy; human; secretory protein; membrane proteins; cancer;
XX inflammatory disease; osteoporosis; neurological disease; gene; ss.
OS Homo sapiens.
XX
PN EPI293569-A2.
XX
PD 19-MAR-2003.
XX
PF 21-MAR-2002; 2002BP-00006586.
XX
PR 14-SEP-2001; 2001JP-00328381.
PR 24-JAN-2002; 2002US-0350435P.
XX
PA (HELI-) HELIX RES INST.
PA (REAS-) RES ASSOC BIOTECHNOLOGY.
XX
PI Iocagui T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S;
PI Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I;
PI Seki N, Yoshikawa T, Otsuka M, Nagahara K, Masuno Y;
XX
DR WPI; 2003-395539/38.
DR P-PDB; ADA55145.
XX
PT New polynucleotides encoding full-length polypeptides, e.g. secretory
PT and/or membrane proteins, useful for developing medicines for diseases in
PT which the gene is involved, or as target molecules for gene therapy.
XX
PS Claim 1; SEQ ID NO 1074; 205pp; English.
XX
CC The present invention relates to novel human secretory or membrane
CC proteins (ADA54072-ADA55710) and their coding sequences (ADA52433-
CC ADA54071). The coding sequences are useful in the gene therapy of
CC diseases caused by abnormalities of the proteins, e.g. cancer,
CC inflammatory diseases, osteoporosis or neurological disease.
XX
SQ Sequence 1988 BP; 519 A; 504 C; 584 G; 381 T; 0 U; 0 Other;

```

Alignment Scores:



XX 04-NOV-2004 (first entry)  
 XX Full length human cDNA useful for treating neurological disease Seq 952.  
 DE  
 XX gene; ss; human; oligo-capping method; diagnostic marker; gene therapy;  
 KM osteoporosis; neurological disease; Alzheimer's disease;  
 KM Parkinson's disease; dementia; short memory; cancer;  
 KM sense or motor function; emotional reaction; fear response; panic;  
 KM osteopathic; neuroprotective; nootropic; antiparkinsonian; cyostatic;  
 KM tranquilliser.  
 XX Homo sapiens.  
 OS  
 XX EP1447413-A2.  
 PN  
 XX 18-AUG-2004.  
 PD  
 XX 12-FEB-2004; 2004EP-00003145.  
 XX  
 XX 14-FEB-2003; 2003JP-00102207.  
 PR  
 XX 09-MAY-2003; 2003JP-00131452.  
 PR  
 XX (REAS-) RES ASSOC BIOTECHNOLOGY.  
 PA  
 XX Ilogai T, Yamamoto J, Nishikawa T, Isono Y, Sugiyama T, Otsuki T;  
 PI Wakamatsu A, Ishii S, Nagai K, Irie R,  
 DR WPI; 2004-583265/57.  
 DR P-PSDB; ADR09402.  
 XX  
 XX New 1995 cDNA, useful for treating osteoporosis, neurological diseases,  
 PT Alzheimer's diseases, Parkinson's diseases, dementia and various cancers.  
 XX  
 XX Claim 1; SEQ ID NO 952; 2686bp; English.  
 PS  
 XX This invention relates to novel, isolated full length human cDNA  
 CC molecules and the encoded proteins thereof. Specifically, it refers to  
 CC cDNA clones obtained by an oligo-capping method, where none of these  
 CC clones are identical to any known human mRNAs. The present invention  
 CC describes an immunoassay to identify agonists and antagonists, as well as  
 CC antibodies, antisense molecules and siRNAs that can all be used to bind  
 CC to and modulate expression of the cDNA molecules. As such, these  
 CC molecules are useful for diagnostic markers or therapeutic targets for  
 CC the various diseases or morbid states. In particular, they are useful in  
 CC gene therapy for treating osteoporosis, neurological disease, Alzheimer's  
 CC disease, Parkinson's disease, dementia, short memory and various cancers,  
 CC as well as for maintaining equilibrium of sense or motor function, and  
 CC for treating emotional reaction, fear response and panic. Accordingly,  
 CC they exhibit osteopathic, neuroprotective, nootropic, antiparkinsonian,  
 CC cyostatic and tranquilliser activities. This polynucleotide is a full  
 CC length human cDNA sequence of the invention. NOTE: This sequence is not  
 CC given in the sequence listing of the specification but can be obtained on  
 CC CD-ROM from the European Patent Office, Vienna Sub-office.  
 XX  
 XX Sequence 2026 BP, 540 A; 513 C; 590 G; 383 T; 0 U; 0 Other;  
 SQ  
 Alignment Scores:  
 Pred. No.: 4, 71e-102 Length: 2026  
 Score: 3210.00 Matches: 590  
 Percent Similarity: 90.40% Conservative: 3  
 Best Local Similarity: 89.94% Mismatches: 11  
 Query Match: 88.36% Indels: 52  
 DB: 13 Gaps: 4  
 US-10-791-017A-2 (1-656) X ADR07446 (1-2026)  
 QY 1 MetAlaSerThrIlePheTyrSerGlnAlaAlaGlnGlnGlyTyrSerAla 20  
 DB 38 ATGGCGCCACGAGATTACAGTACTTATGACCAAGCTGACGCGAGGCTCAGGCT 97  
 QY 21 TTTThAlaGlnProThrGlnGlyTyrAlaGlnThrGlnAlaTyrGlyGlnGlnSer 40

DB 98 TACACGCCACGCCCACTCAAGATATGCACAGCACCCAGGATATGGCAACAAGC 157  
 QY 441 TTTGlyThrTyrGlnGlnProThrAspValSerTyrThrGlnAlaGlnThrAlaThr 60  
 DB 158 TATGAACTATGAGACAGCCCACTGATGTCAGCTATACCCAGGCTCAGACCACTGAAAC 217  
 QY 61 TTTGlyGlnThrAlaTyrAlaThrSerTyrGlyGlnProThrGlyTyrThrThrPro 80  
 DB 218 TATGGGACAGCCGCTATGCACTCTTATGACAGCTCCCACTGGTTTATCACTCCA 277  
 QY 81 ThrAlaProGlnAlaTyrSerGlnProValGlnGlyThrGlyAlaTyrAspThr 100  
 DB 278 TCTACCTCTCTAC-----ACCTATCTCT 304  
 QY 101 ThrThrAlaThrValThrThrGlnAlaSerTyrAlaAlaGlnSerAlaTyrGlyThr 120  
 DB 305 ACACAGCCGACT-----AGTTATGATCAGAGAGCT----- 334  
 QY 121 GlnProAlaTyrProAlaTyrGlyGlnGlnProAlaAlaThrAlaProThrAspProGln 140  
 DB 335 -----TACTCTCAGACGACGACGACGACCTGACCTCAAGACCGCAG 376  
 QY 141 AspGlyAsnLysProThrGlnThrSerGlnProGlnSerSerThrGlyTyrAsnGln 160  
 DB 377 GATGAAACCAAGCCCACTGAGACTAGTCACTCACTGACAGGGGTTACACAG 436  
 QY 161 ProSerLeuGlyTyrGlyGlnSerAsnTyrSerTyrProGlnValProGlySerTyrPro 180  
 DB 437 CCCAGCTAGATATGAGACAGAGTAACACTACAGTTATCCCAAGGTACCTGGAGAGTACCC 496  
 QY 181 MetGlnProValThrAlaProProSerTyrProProThrSerTyrSerSerThrGlnPro 200  
 DB 497 ATGACAGCACTGACTGACCTCACTCACTCACTCACTCACTCACTCACTCACTCACTCACT 535  
 QY 201 ThrSerTyrAspGlnSerSerTyrSerGlnGlnAlaThrTyrGlyGlnProSerSerTyr 220  
 DB 536 -----AGCTAT 541  
 QY 221 GlyGlnGlnSerSerTyrGlyGlnGlnSerSerTyrGlyGlnGlnProProThrSerTyr 240  
 DB 542 GACAGAGAGTATGATGATGATCAACAAGCAGCTATGAGCAGCAGCTCCCACTAGTTAC 601  
 QY 241 ProProGlnThrTyrSerTyrSerGlnAlaProSerGlnTyrSerGlnGlnSerSerSer 260  
 DB 602 CCACCCCAACTGATCTCAACACCAAGCTCCAAATGATGATGATGATGATGATGATGATGAT 661  
 QY 261 TTTGlyGlnGlnSerSerPheArgGlnAspHisProSerSerMetGlyValTyrGlyGln 280  
 DB 662 TACGGGACAGAGCTTCACTTCCAGACAGACACCCCACTGATGATGATGATGATGATGATGAT 721  
 QY 281 GlnSerGlyGlyPheSerGlyProGlyGlnAsnArgSerMetSerGlyProAspAsnArg 300  
 DB 722 GATCTGAGAGATTTTCCGAGACAGAGAGAACCGAGGATGATGATGATGATGATGATGATGAT 781  
 QY 301 GlyArgGlyArgGlyGlyPheAspArgGlyGlyMetSerArgGlyGlyArgGlyGlyGly 320  
 DB 782 GCGAGGGGAGAGAGAGGAGATTTGATCTGAGGATGAGGATGAGGATGAGGATGAGGATGAGGAT 841  
 QY 321 ArgGlyGlyMetGlySerAlaGlyGlyArgGlyGlyPheAsnLysProGlyGlyProMet 340  
 DB 842 CCGCGTGGATATGGCAGCGCTGAG 901  
 QY 341 AspGlnGlyProAspLeuAspLeuGlyProProValAspProAspGlnAspSerAspAsn 360  
 DB 902 GATGAAGAGACCAATCTGATCTAGGCCCACTGTGATGATCAATGATGATGATGATGATGATGAT 961  
 QY 361 SerAlaIleTyrAlaGlnGlyLeuAsnAspSerValThrLeuAspAspLeuAlaAspPhe 380  
 DB 962 AGTGAATATTTATGACAGATTAATGACAGGTGATCTTATATGATATGATGATGATGATGATGAT 1021  
 QY 381 PheLysGlnCysGlyValValLysMetAsnLysArgThrGlyGlnProMetIleHisIle 400  
 DB 1022 TTTTACAGTGTGGGTTTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1081





**This Page Blank (usptn)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

## OM protein - nucleic search, using frame\_plus.p2n model

Run on: February 20, 2005, 22:00:21 ; Search time 291.042 Seconds  
(without alignments)  
3688.116 Million cell updates/sec

Title: US-10-791-017A-2  
Perfect score: 3633  
Sequence: 1 MASTDSTYSQAAAGGYSK.....GGPKKDKGHRGRRDRPV 656

Scoring table:  
BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 1202784 seqs, 818138359 residues  
Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

## Command line parameters:

-MODE=frame+ p2n.model -DEV=x1p  
-Q=/cgn2\_1/USPTO.spool\_p/US10791017/runat.17022005.125808.22107/app.query.faetaa\_1.1358  
-DB=Issued Patents NA -QEMT=fastap -SUFFIX=p2n.rnt -MINMATCH=0.1 -LOOPL=0  
-LOOPEXT=0 -INITs=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human0.cdt  
-LIST=45 -DOCLINK=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTWTA=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US10791017@cgn2\_1.141@runat.17022005.125808.22107 -NCPU=6 -ICPU=3  
-NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPELCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -MARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

## Database :

Issued Patents NA.\*  
1: /cgn2\_6/ptodata/1/lna/5A COMB.seq.\*  
2: /cgn2\_6/ptodata/1/lna/5B COMB.seq.\*  
3: /cgn2\_6/ptodata/1/lna/6A COMB.seq.\*  
4: /cgn2\_6/ptodata/1/lna/6B COMB.seq.\*  
5: /cgn2\_6/ptodata/1/lna/PCTUS COMB.seq.\*  
6: /cgn2\_6/ptodata/1/lna/backfile1.seq.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3633	100.0	2371	2	US-08-343-443B-1
2	3233	89.0	6002	4	US-09-949-016-13636
3	3178.5	87.5	1785	4	US-09-949-016-5043
4	3178.5	87.5	1785	4	US-09-949-016-5044
5	2826.5	77.8	1783	4	US-09-949-016-1954
6	1420	39.1	2432	1	US-08-437-027-18
7	1273.5	35.1	1939	4	US-09-919-039-322
8	958	26.4	35784	4	US-09-949-016-16785
9	958	26.4	35784	4	US-09-949-016-16786
10	676	18.6	454	4	US-09-513-999C-1657
11	609	16.8	411	4	US-09-621-976-13361
12	566	15.6	601	4	US-09-949-016-66382

13	486	13.4	1682	4	US-09-320-132-82	Sequence 82, Appl
14	484	13.3	954	2	US-08-343-443B-106	Sequence 106, App
15	484	13.3	1684	4	US-09-919-039-323	Sequence 323, App
16	471	13.0	333	4	US-09-513-999C-12062	Sequence 12062, A
17	405	11.1	601	4	US-09-949-016-176641	Sequence 176641,
18	405	11.1	601	4	US-09-949-016-176641	Sequence 176641,
19	397.5	10.9	450	3	US-09-370-838-145	Sequence 145, App
20	397.5	10.9	450	3	US-09-854-133-145	Sequence 145, App
21	386.5	10.6	3147	2	US-08-781-802-7	Sequence 7, Appl
22	386.5	10.6	3147	3	US-08-694-078-7	Sequence 7, Appl
23	386.5	10.6	3147	3	US-09-058-260-7	Sequence 7, Appl
24	368.5	10.1	3600	1	US-08-537-002A-5	Sequence 5, Appl
25	368.5	10.1	3600	3	US-08-863-010-5	Sequence 5, Appl
26	368.5	10.1	3600	3	US-09-024-429-5	Sequence 5, Appl
27	360.5	9.9	2830	2	US-09-010-928B-1	Sequence 1, Appl
28	353.5	9.7	6109	4	US-09-795-061-1	Sequence 1, Appl
29	351.5	9.7	1926	3	US-09-249-585A-4	Sequence 4, Appl
30	351.5	9.7	1931	2	US-09-130-114-2	Sequence 2, Appl
31	347	9.6	4411529	3	US-09-103-840A-1	Sequence 1, Appl
32	345	9.5	4411529	3	US-09-103-840A-1	Sequence 1, Appl
33	344	9.5	2082	4	US-09-818-780-67	Sequence 67, Appl
34	343	9.4	4403765	3	US-09-103-840A-2	Sequence 2, Appl
35	341.5	9.4	4403765	3	US-09-103-840A-2	Sequence 2, Appl
36	341	9.4	2824	2	US-09-010-928B-3	Sequence 3, Appl
37	337.5	9.3	4409	4	US-09-331-347C-22	Sequence 22, Appl
38	331	9.1	6200	4	US-09-795-061-3	Sequence 3, Appl
39	330	9.1	6158	4	US-09-919-497-6	Sequence 6, Appl
40	329.5	9.1	3171	4	US-09-169-768-15	Sequence 15, Appl
41	329.5	9.1	3181	1	US-08-655-086-1	Sequence 1, Appl
42	329.5	9.1	3349	4	US-09-169-768-13	Sequence 13, Appl
43	329.5	9.1	3541	4	US-09-169-768-9	Sequence 9, Appl
44	329	9.1	2493	1	US-07-977-434-5	Sequence 5, Appl
45	329	9.1	2493	1	US-08-458-819-5	Sequence 5, Appl

## ALIGNMENTS

RESULT 1  
US-08-343-443B-1  
Sequence 1, Application US/08343443B  
Patent No. 5968734  
GENERAL INFORMATION:  
APPLICANT: Aurias, Alain  
APPLICANT: Delattre, Olivier  
APPLICANT: Desmarte, Chantal  
APPLICANT: Meiot, Thomas  
APPLICANT: Peter, Martine  
APPLICANT: Plooungastel, Beatrice  
APPLICANT: Thomas, Gilles  
APPLICANT: Zucman, Verónica  
TITLE OF INVENTION: NUCLEIC ACID CORRESPONDING TO A GENE OF  
TITLE OF INVENTION: CHROMOSOME 22 INVOLVED IN RECURRENT CHROMOSOMAL  
TITLE OF INVENTION: TRANSLATIONS ASSOCIATED WITH THE DEVELOPMENT OF CANCEROUS  
TITLE OF INVENTION: TUMORS, AND NUCLEIC ACIDS OF FUSION RESULTING FROM SAID  
TITLE OF INVENTION: TRANSLOCATIONS  
NUMBER OF SEQUENCES: 129  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Weiser & Associates  
STREET: 230 South Fifteenth Street  
CITY: Philadelphia  
STATE: PA  
COUNTRY: USA  
ZIP: 19102  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: AEDIT 1.0 DOS text editor  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/343, 443B  
FILING DATE: 18-NOV-1994  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT/FR93/00494  
 FILING DATE: 19-MAY-1993  
 PRIOR APPLICATION DATA: FR 92/06123  
 APPLICATION NUMBER: 989.6121P  
 FILING DATE: 20-MAY-1992  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Weiser, Gerard J.  
 REGISTRATION NUMBER: 19,763  
 REFERENCE/DOCKET NUMBER: 989.6121P  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 215-875-8383  
 TELEFAX: 215-875-8394  
 INFORMATION FOR SEQ ID NO: 1:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 2371 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: double  
 TOPOLOGY: linear  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: 25..1992  
 US-08-343-443B-1  
 Alignment Scores:  
 Pred. No.: 4.3e-186 Length: 2371  
 Score: 3633.00 Matches: 656  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 Gaps: 0  
 US-10-791-017A-2 (1-656) x US-08-343-443B-1 (1-2371)

201 ThrSerTyrAspGlnSerSerTyrSerGlnGlnAenThrTyrGlyGlnProSerSerTyr 220  
 625 ACTAGTTATGATCAGAGCAGTTACTCTGACGAGAACCTTATGGGACCGAGCACTAT 684  
 221 GtGlnGlnSerSerTyrGlyGlnGlnSerSerTyrGlyGlnGlnProProThrSerTyr 240  
 685 GGAACAGCAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 744  
 241 ProProGlnThrGlySerTyrSerGlnAProSerGlnTyrSerGlnGlnGlnSerSer 260  
 745 CCACCCCAACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 804  
 261 TyrGlyGlnGlnSerSerPheArgGlnAProSerSerMetGlyValTyrGlyGln 280  
 805 TACGGCAGCAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 864  
 281 GluSerGlyGlyPheSerGlyProGlyGlnAunArgSerMetSerGlyProAAspAunArg 300  
 865 GAGTCTGAGAGATTTTCCGAGCAGAGAGAAACCGAGCATGATGATGATGATGATGATGAT 924  
 301 GlyArgGlyValArgGlyPheAAspArgGlyGlyMetSerArgGlyGlyValArgGlyGly 320  
 925 GGCAGGGGAGAGAGGGGATTTGATGATGATGATGATGATGATGATGATGATGATGATGAT 984  
 321 ArgGlyGlyMetGlySerAglGlyAglGlyGlyGlyPheAAspArgProGlyGlyPheMet 340  
 985 CCGGGTGAATGGGACGGCTCGAGAGCCAGAGTGGCTTCAATAGCTGGTGGACCAATG 1044  
 341 AspGlyGlyProAAspLeuAAspLeuGlyProProValAAspProAAspGlyAAspSerAAspAun 360  
 1045 GATGAAGAGACCAAGATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1104  
 361 SerAAlaIleTyrValGlnGlyLeuAAspSerValThrLeuAAspLeuAAspPhe 380  
 1105 AGTCGATTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1164  
 381 PheLeuGlnCysGlyValValValValValValValValValValValValValValVal 400  
 1165 TTTTAAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1224  
 401 TyrLeuAAspLeuGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 420  
 1225 TACCTGAGACAGAGAAACAGAGAAACCCAAAGGAGATGATGATGATGATGATGATGAT 1284  
 421 ProThrAlaValAlaValAlaValAlaValAlaValAlaValAlaValAlaValAlaVal 440  
 1285 CCACCTGAGCAGAGCTCCGAGATGATGATGATGATGATGATGATGATGATGATGATGAT 1344  
 441 LysValSerLeuAlaArgLysLysProPheMetAAspSerMetArgGlyGlyLeuProPro 460  
 1345 AAGGTCTCTGCTGCTGAGAAAGCTCCAGATGAAACATGATGATGATGATGATGATGAT 1404  
 461 ArgGlnGlyValArgLysMetProProPheAAspArgGlyGlyProGlyGlyPro 480  
 1405 CGTGAGAGGAGAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1464  
 481 GlyGlyProMetGlyValArgMetGlyValArgGlyGlyValArgGlyGlyLysPheProProArg 500  
 1465 GGGGAGACCAATGGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1524  
 501 GlyProArgGlySerArgGlyAAspProSerGlyGlyValAAspValGlnAAspArgGly 520  
 1525 GAGCCCGGGGTTCCGAGAGAAACCTCTGAGAGAGAAACCTCCAGACCGAGGCTGGA 1584  
 521 AspTyrGlnCysPheAAspProGlyCysGlyAAspGlnAAspPheAAspArgGlyGlyCys 540  
 1585 GACTGGCAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1644  
 541 AAspGlyCysValAlaProLysProGlyGlyPheLeuProProProPheProProProGly 560  
 1645 AACCAAGTGAAGGACCCCAAGGCTGGAAGGCTTCTCCGACACCTTCTCCGCGGGGT 1704  
 561 GlyAAspArgGlyArgGlyGlyProGlyGlyMetArgGlyGlyValArgGlyGlyLeuMetAAsp 580

Db 1705 GGATGATGCTGGCAGAGGTGGCCCTGCTGGCATGGGGGAGAAAGGTGGCTTATGAT 1764  
Qy 581 ATGGTGTGCTGGCAGAGGTGGCCCTGCTGGCATGGGGGAGAAAGGTGGCTTATGAT 1764  
Db 1765 CGTGTGTGCTGGCAGAGGTGGCCCTGCTGGCATGGGGGAGAAAGGTGGCTTATGAT 1824  
Qy 601 GATGTGTGCTGGCAGAGGTGGCCCTGCTGGCATGGGGGAGAAAGGTGGCTTATGAT 1824  
Db 1825 GGTGTGTGCTGGCAGAGGTGGCCCTGCTGGCATGGGGGAGAAAGGTGGCTTATGAT 1884  
Qy 621 GGTGTGTGCTGGCAGAGGTGGCCCTGCTGGCATGGGGGAGAAAGGTGGCTTATGAT 1884  
Db 1885 GGTGTGTGCTGGCAGAGGTGGCCCTGCTGGCATGGGGGAGAAAGGTGGCTTATGAT 1944  
Qy 641 GGTGTGTGCTGGCAGAGGTGGCCCTGCTGGCATGGGGGAGAAAGGTGGCTTATGAT 1944  
Db 1945 GGTGTGTGCTGGCAGAGGTGGCCCTGCTGGCATGGGGGAGAAAGGTGGCTTATGAT 1992  
RESULT 2  
US-09-949-016-13696  
/ Sequence 13696, Application US/09949016  
/ Patent No. 6812339  
/ GENERAL INFORMATION:  
/ APPLICANT: VENTER, J. Craig et al.  
/ TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
/ WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
/ FILE REFERENCE: CL001307  
/ CURRENT APPLICATION NUMBER: US/09/949, 016  
/ PRIOR FILING DATE: 2000-04-14  
/ PRIOR APPLICATION NUMBER: 60/241,755  
/ PRIOR FILING DATE: 2000-10-20  
/ PRIOR APPLICATION NUMBER: 60/237,768  
/ PRIOR FILING DATE: 2000-10-03  
/ PRIOR APPLICATION NUMBER: 60/231,498  
/ PRIOR FILING DATE: 2000-09-08  
/ NUMBER OF SEQ. ID NOS: 207012  
/ SOFTWARE: FASTSEQ for Windows Version 4.0  
/ SEQ ID NO 13696  
/ LENGTH: 6002  
/ TYPE: DNA  
/ ORGANISM: Human  
US-09-949-016-13696  
Alignment Scores:  
Pred. No.: 2,398-164 Length: 6002  
Score: 3233.00 Matches: 601  
Percent Similarity: 93.60% Conservative: 13  
Best Local Similarity: 91.62% Mismatches: 42  
Query Match: 88.99% Indels: 2  
DB: 4 Gaps: 0  
US-10-791-017a-2 (1-656) x US-09-949-016-13696 (1-6002)  
Qy 1 MetAlaSerThrAspTyrSerThrTyrSerGlnAlaAlaGlnGlnGlnTyrSerAla 20  
Db 2034 ATGGCATTCATGATTAACATTAACATTAACATTAACATTAACATTAACATTAACATTA 2093  
Qy 21 TTTTAAAGTGTGCTGGCAGAGGTGGCCCTGCTGGCATGGGGGAGAAAGGTGGCTTATGAT 2153  
Db 2094 TTTTAAAGTGTGCTGGCAGAGGTGGCCCTGCTGGCATGGGGGAGAAAGGTGGCTTATGAT 2153  
Qy 41 TTTTAAAGTGTGCTGGCAGAGGTGGCCCTGCTGGCATGGGGGAGAAAGGTGGCTTATGAT 2213  
Db 2154 TTTTAAAGTGTGCTGGCAGAGGTGGCCCTGCTGGCATGGGGGAGAAAGGTGGCTTATGAT 2273  
Qy 61 TTTTAAAGTGTGCTGGCAGAGGTGGCCCTGCTGGCATGGGGGAGAAAGGTGGCTTATGAT 2273  
Db 2274 TTTTAAAGTGTGCTGGCAGAGGTGGCCCTGCTGGCATGGGGGAGAAAGGTGGCTTATGAT 2333

Qy 101 ThrThrAlaThrValThrThrThrGlnAlaSerTyrAlaAlaGlnSerAlaTyrGlyThr 120  
Db 2334 ACCATGCTTACAGTCAAC 2393  
Qy 121 GlnProAlaTyrProAlaTyrGlyGlnGlnProAlaAlaThrAlaProThrArgProGln 140  
Db 2394 CAGCTGCTTATTCAGCTTATTCAGCTTATTCAGCTTATTCAGCTTATTCAGCTTATTCAG 2453  
Qy 141 AspGlyAsnLysProThrGlnThrSerGlnProGlnSerSerThrGlyTyrAsnGln 160  
Db 2454 AATGAAACCAAGCCCACTGAGTCACTCACTCACTCACTCACTCACTCACTCACTCACTCA 2513  
Qy 161 ProSerLeuGlyTyrGlyGlnSerAsnTyrSerTyrProGlnValProGlySerTyrPro 180  
Db 2514 CCAAGCTTACAGTCAAC 2573  
Qy 181 MetGlnProValThrAlaProProSerTyrProProThrSerTyrSerSerThrGlnPro 200  
Db 2574 ATGACGCTTACAGTCAAC 2633  
Qy 201 ThrSerTyrAsnGlnSerSerTyrSerGlnGlnAsnThrTyrGlyGlnProSerSerTyr 220  
Db 2634 ACTAGTTATGATCAGAGCAGTATTAATCTCACTCACTCACTCACTCACTCACTCACTCA 2693  
Qy 221 GATGCTGTGCTGGCAGAGGTGGCCCTGCTGGCATGGGGGAGAAAGGTGGCTTATGAT 2753  
Db 2694 GATGCTGTGCTGGCAGAGGTGGCCCTGCTGGCATGGGGGAGAAAGGTGGCTTATGAT 2813  
Qy 241 ProProGlnThrGlySerTyrSerGlnAlaProSerGlnThrSerGlnSerSerSer 260  
Db 2754 CCAAGCTTACAGTCAAC 2813  
Qy 261 TTTTAAAGTGTGCTGGCAGAGGTGGCCCTGCTGGCATGGGGGAGAAAGGTGGCTTATGAT 2873  
Db 2814 TTTTAAAGTGTGCTGGCAGAGGTGGCCCTGCTGGCATGGGGGAGAAAGGTGGCTTATGAT 2933  
Qy 281 GluSerGlyGlyPheSerGlyProGlyGlnAsnArgSerSerSerGlyProAspAsnArg 300  
Db 2874 GATGCTGTGCTGGCAGAGGTGGCCCTGCTGGCATGGGGGAGAAAGGTGGCTTATGAT 3063  
Qy 301 GATGCTGTGCTGGCAGAGGTGGCCCTGCTGGCATGGGGGAGAAAGGTGGCTTATGAT 3113  
Db 2934 GATGCTGTGCTGGCAGAGGTGGCCCTGCTGGCATGGGGGAGAAAGGTGGCTTATGAT 3173  
Qy 321 ArgGlyGlyMetGlySerAlaGlyGlnArgGlyGlyPheAsnLysProGlyGlyProMet 340  
Db 2994 TGGGCTGTGCTGGCAGAGGTGGCCCTGCTGGCATGGGGGAGAAAGGTGGCTTATGAT 3463  
Qy 341 AspGlyGlyProAspLeuAspLeuGlyProProValAspProAspGlnAspSerAspAsn 360  
Db 3054 GATGCTGTGCTGGCAGAGGTGGCCCTGCTGGCATGGGGGAGAAAGGTGGCTTATGAT 3713  
Qy 361 SerAlaIleTyrValGlnGlyLeuAsnAspSerValThrLeuAspLeuAlaAspPhe 380  
Db 3114 AGTCAATTTATTAACAGATTAACAGATTAACAGATTAACAGATTAACAGATTAACAG 3933  
Qy 381 PheLeuGlnGlyValValLysMetAsnLysArgThrGlyGlnProMetIleHisIle 400  
Db 3174 TTTTAAAGTGTGCTGGCAGAGGTGGCCCTGCTGGCATGGGGGAGAAAGGTGGCTTATGAT 4063  
Qy 401 TTTTAAAGTGTGCTGGCAGAGGTGGCCCTGCTGGCATGGGGGAGAAAGGTGGCTTATGAT 420  
Db 3234 TTTTAAAGTGTGCTGGCAGAGGTGGCCCTGCTGGCATGGGGGAGAAAGGTGGCTTATGAT 4273  
Qy 421 ProThrAlaValAlaAlaValAlaGlnTyrPheAspGlyLysAspPheGlnGlySerLys 440  
Db 3294 CCAAGCTTACAGTCAAC 4463  
Qy 441 LysValSerLeuAlaArgLysLysProProMetAsnSerMetArgGlyGlyLeuProPro 460  
Db 3354 AATGCTTCTTCTGCTGGCAGAGGTGGCCCTGCTGGCATGGGGGAGAAAGGTGGCTTATGAT 4713

QY 461 ArgGluGlyArgGlyMetProProProLeuArgGlyGlyProGlyGlyProGlyGlyPro 480  
 Db 3414 CGTGGAGGCGAGAGGAGTCCACCACTCTGCGAGAGTCCAGAGGCCAGAGAGTCT 3473  
 QY 481 G1YGLYProMetGlyArgMetGlyGlyArgGlyGlyArgGlyGlyPheProProArg 500  
 Db 3474 GGGGGACCCATGGGTTCATCATGGAGGCGCTGAGAGATAGAGAGGCTCCCTCCAGA 3533  
 QY 501 G1YProArgGlySerArgGlyArgProSerGlyGlyGlyArgGlyGlyGlyGlyGly 520  
 Db 3534 GAGACCCAGGAGTCCCGAGGAGAACCTCTGAGAGAGAGAACCTCCAGCACCAAGCTGA 3593  
 QY 521 AspTrpGlnCysProAsnProGlyCysGlyAsnGlnAsnPheAlaTrpArgTrpGly 540  
 Db 3594 GACAGGAGTCTCCCATTCGGGTGTCGAAACCAACCTTCGCTGAGAACAGAGAGC 3653  
 QY 541 AsnGlnCysAlaProLysProGlyGlyPheLeuProProPheProProProGly 560  
 Db 3654 AACAAAGTAAAGGCTCAAGAGCTGAGAGCTTCTCCCGCACCCCTCCAGCCCGGGT 3713  
 QY 561 G1YAspArgGlyArgGlyGlyProGlyGlyMetArgGlyGlyArgGlyGlyLeuMetAsp 580  
 Db 3714 GGTGATATGAGCAAGGTGGCTGTGTGCAATGGAGAGAGAGGTGGCTCATAGAT 3773  
 QY 581 ArgGlyGlyProGlyGlyMetPheArgGlyGlyArgGlyGlyAspArgGlyGlyPheArg 600  
 Db 3774 CATGTGTCTCCGCTGAGAGTTCAGAGTGTGCTGTGTAGAGACAGAAAGTGGCTTCT 3833  
 QY 601 G1YGLYArgGlyMetAspArgGlyGlyPheGlyGlyGlyArgGlyGlyProGlyGly 620  
 Db 3834 GGTGGCTG-GGACATGAGCCAGAGGTGGCTTGTGTGAGAGAGACAAAGTGGCTGGGG 3892  
 QY 621 ProProGlyProLeuMetGlnMetGlyGlyArgGlyGlyArgGlyGlyProGly 640  
 Db 3893 CCCCCGGA-CCTTGATGATGAACCAATGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3951  
 QY 641 LysMetAspLysGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 656  
 Db 3952 AAAACGATAAAGCGAGCACTGTCAAGAGCGAGAGATCAAGCTTAC 3999

RESULT 3  
 US-09-949-016-5043  
 ; Sequence 5043, Application US/09949016  
 ; Patent No. 6812339  
 ; GENERAL INFORMATION:  
 ; APPLICANT: VENTER, J. Craig et al.  
 ; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
 ; FILE REFERENCE: CU001307  
 ; CURRENT APPLICATION NUMBER: US/09/949,016  
 ; CURRENT FILING DATE: 2000-04-14  
 ; PRIOR APPLICATION NUMBER: 60/241,755  
 ; PRIOR FILING DATE: 2000-10-20  
 ; PRIOR APPLICATION NUMBER: 60/237,768  
 ; PRIOR FILING DATE: 2000-10-03  
 ; PRIOR APPLICATION NUMBER: 60/231,498  
 ; PRIOR FILING DATE: 2000-09-08  
 ; NUMBER OF SEQ ID NOS: 207012  
 ; SOFTWARE: FASTSEQ for Windows Version 4.0  
 ; SEQ ID NO 5043  
 ; LENGTH: 1785  
 ; TYPE: DNA  
 ; ORGANISM: Human  
 ; US-09-949-016-5043

Alignment Scores:  
 Pred. No.: 6,89e-162 Length: 1785  
 Score: 3178.50 Matches: 582  
 Percent Similarity: 88.72% Conservative: 0  
 Best Local Similarity: 88.72% Mismatches: 1  
 Query Match: 87.49% Indels: 73  
 Gaps: 1

US-10-791-017a-2 (1-656) x US-09-949-016-5043 (1-1785)  
 QY 1 MetAlaSerThrAspTrpSerThrTrpSerGlnAlaAlaGlnGlnGlyTrpSerAla 20  
 Db 34 ATGGCTTCAAGGATTAAGTAACTTAATACCAAGCTGACAGGCGAGAGGCTTCAAGTGT 93  
 QY 21 TyrThrAlaGlnProThrGlnGlyGlyAlaGlnThrThrGlnAlaTrpGlyGlnGlnSer 40  
 Db 94 TACACCGCCAGGCCACTAGAGATATGACAGACCAACCAAGCATATGGGCAACAAAGC 153  
 QY 41 TyrGlyThrThrGlyGlnProThrAspValSerTrpThrGlnAlaGlnThrThrAlaThr 60  
 Db 154 TATGAACTATAGACAGGCCACTGATGTACGATATACCAAGCTGACCACTGCAACC 213  
 QY 61 TyrGlyGlnThrAlaThrAlaThrSerTrpGlyGlnProProThrGlnGlyTrpThrPro 80  
 Db 214 TATGGCAAGCCGCTATACACTTTATGAGACAGCTCCACTGATTAATCTACCA 273  
 QY 81 ThrAlaProGlnAlaTrpSerGlnProValGlnGlyTrpGlyThrGlyAlaTrpAspThr 100  
 Db 274 ACTGCCCCAGGCAATACAGCAGCAGCTGTCCAGGGTATGGCACTGTGTATATAC 333  
 QY 101 ThrThrAlaThrValThrThrThrGlnAlaSerTrpAlaAlaGlnSerAlaTrpGlyThr 120  
 Db 334 ACCACTGTACAGTCAACCAACCAAGGCTCTATGACGTCACTGATATGACACT 393  
 QY 121 GlnProAlaThrProAlaThrGlyGlnGlnProAlaAlaThrAlaProThrArgProGln 140  
 Db 394 CAGCTGCTTATTCACCTATAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 453  
 QY 141 AspGlyAsnLysProThrGlnThrSerGlnProGlnSerSerThrGlyGlyTrpAsnGln 160  
 Db 454 GATGAAACAGGCCACTAGACTACTCAACTCATCTACAGACAGGGGTTTCAACAG 513  
 QY 161 ProSerLeuGlyTrpGlyGlnSerAsnTrpSerTrpProGlnValProGlySerTrpPro 180  
 Db 514 CCAAGCTAGATATAGACAGAGTAACTACATTAATCCAGGATACCTGGAGACTACCCC 573  
 QY 181 MetGlnProValThrAlaProProSerTrpProProThrSerTrpSerSerThrGlnPro 200  
 Db 574 ATGACAGCAATCACTGACCTCATCTTACCTTACCAAGCATATCTCTTACACAGCCG 633  
 QY 201 ThrSerTrpAspGlnSerSerTrpSerGlnGlnAsnThrTrpGlyGlnProSerSerTrp 220  
 Db 634 ACTGATTAATGATACAGCACTTACTCTACAGAGAACTTAATGAGCAACGAGACACTAT 653  
 QY 221 GylGlnGlnSerSerTrpGlyGlnGlnSerSerTrpGlyGlnGlnProProThrSerTrp 240  
 Db 694 GAGACAGAGTATGATGATGATCAACAAAGCAAGTATAGGAGAGGCTCCCACTAGTAC 753  
 QY 241 ProProGlnThrGlySerTrpSerGlnAlaProSerGlnTrpSerGlnGlnSerSerSer 260  
 Db 754 CCAACCCCAAGTGAATCTTACAGCCAGGCTCCAAAGTATATGCAACAGAGACAGCAGC 813  
 QY 261 TyrGlyGlnGlnSerSerPheArgGlnAspHisProSerSerMetGlyValTrpGlyGln 280  
 Db 814 TACGGAG 828  
 QY 281 GluSerGlyGlyPheSerGlyProGlyGlnAsnArgSerMetSerGlyProAspAsnArg 300  
 Db 828 ----- 828  
 QY 301 GlyArgGlyArgGlyGlyPheAspArgGlyGlyMetSerArgGlyGlyArgGlyGly 320  
 Db 828 ----- 828  
 QY 321 ArgGlyGlyMetGlySerAlaGlyArgGlyGlyPheAsnLysProGlyGlyProMet 340  
 Db 829 -----CCCATG 834  
 QY 341 AspGluGlyProAspLeuAspLeuGlyProProValAspProAspGlyLysAspAspAsn 360  
 Db 835 GATGAGAGACAGATTTGATCTAGGCCCCAGCTGTGATCTCAATGATAGACTCTGACAAAC 894

```

QY 361 SerAlaIleTyrValGlnGlyLeuAsnAspSerValThrLeuAspAspLeuAlaAspPhe 380
Db 895 AGTGCAATTATGATGACAGATTAATAAGCATGTGCTCTAGATGATCTGGCAGACTTC 954
QY 381 PheLeuGlnCyseGlyValValLeuMetAsnLysArgThrGlnProMetIleH1e 400
Db 955 TTTAAGCAGTGTGGGTGTTTAAATGATGACAGAACTGGGCAACCAATGATCAATC 1014
QY 401 TyrLeuAspLysGlyThrGlyLysProLysGlyAspAlaThrValSerTyrGluAspPro 420
Db 1015 TACCTGACAGAGAAACAGAAAGCCCAAGGCATCCACAGTCTCTATGAAGACCA 1074
QY 421 ProThrAlaLysAlaValGlnTyrPheAspGlyLysAspPheGlnGlySerLysLeu 440
Db 1075 CCACCTGCCAAGGCTGCTGGTAATGTTGATGGAAATTTTCAAGGAGCAAACTT 1134
QY 441 LysValSerLeuAlaArgLysLysProMetAsnSerMetArgGlyLysLeuProPro 460
Db 1135 AAGTCTCCTTGTCTCGAAGAAAGCTCCATGAAACGTATGGGGGTGCTTCCACCC 1194
QY 461 ArgGlnGlyValArgLysMetProProProLeuArgGlyLysProGlyLysPro 480
Db 1195 CGTGAAGGACAGAGCATGCCACCACTCCGTGAGAGTCCAGGAGGCCAGAGGCTCT 1254
QY 481 GlnGlyProMetGlyArgMetGlyLysArgGlyLysAspArgGlyLysPheProProArg 500
Db 1255 GGGGAGCCCAAGGATCGCATGGAGGCGCTGGAGAAATAGAGAGGCTTCCCTCCAA 1314
QY 501 GlyProArgLysSerArgGlyAsnProSerGlyLysLysValGlnH1eArgAlaGly 520
Db 1315 GAGCCCGGGGTTCCGAGGGAACCCCTCTGAGAGAAACGTCCAGCACAGAGCTGGA 1374
QY 521 AspTyrGlnCyseProAsnProGlyLysGlyAsnGlnAsnAspAlaTyrArgThrGly 540
Db 1375 GACTGGCAGTGTCCCATCCGGGTGGAAACCAAGAACTTCCGCTGAGAAACAGAGTGC 1434
QY 541 AsnGlnCyseLysAlaProLysProGlnGlyLysPheLeuProProPheProProGly 560
Db 1435 AACCAAGTAAAGCCCAAGGCTTCCCTCCGACCCCTTCCGCCCCCGGGT 1494
QY 561 GlyAspArgLysArgGlyLysProGlyLysMetArgGlyLysArgGlyLysLeuMetAsp 580
Db 1495 GGTGATCTGTGGCAGAGTGGCCCTGTGGCATGGGGAGAGAGAGTGGCTTCATGAT 1554
QY 581 ArgGlyLysProGlyLysMetPheArgGlyLysArgGlyLysAspArgGlyLysPheArg 600
Db 1555 CGTGTGTCTCCCGTGAATGTTCAAGAGTGGCCGTGTGAGACAGAGTGGCTTCGT 1614
QY 601 GlnGlyArgGlyLysMetAspArgGlyLysPheGlnGlyLysArgArgGlyLysProGly 620
Db 1615 GGTGGCGGGGAGTGAACGAGGTGGCTTGTGTGGAGAGACAGAGTGGCTTCGGGGG 1674
QY 621 ProProGlyProLeuMetGlnGlnMetGlyLysArgArgGlyLysArgGlyLysProGly 640
Db 1675 CCCCCCTGACCTTGTATGAAACAGATGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1734
QY 641 LysMetAspLysGlyLysLysArgGlnGlnLysArgArgAspArgProTyr 656
Db 1735 AAAATGATTAAGGCGAGCAACCGTCAAGGAGCGAGAGATCGGCCCTTAC 1782

```

```

RESULT 4
US-09-949-016-5044
; Sequence 5044, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755

```

```

; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO: 5044
; LENGTH: 1785
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-5044

Alignment Scores:
Pred. No.: 6.89e-162 Length: 1785
Score: 3178.50 Matches: 582
Percent Similarity: 88.72% Conservative: 0
Best Local Similarity: 88.72% Mismatches: 1
Query Match: 87.49% Indels: 73
DB: 4 Gaps: 1

US-10-791-017A-2 (1-656) x US-09-949-016-5044 (1-1785)
QY 1 MetAlaSerThrAspTyrSerThrTyrSerGlnAlaAlaGlnGlnGlyTyrSerAla 20
Db 34 ATGGCTTCCAGGATTAACCTTATGACCACTGACGAGGAGGCTTACAGTGTCT 93
QY 21 TyrThrAlaGlnProThrGlnGlyTyrAlaGlnThrThrGlnAlaTyrGlyGlnGlnSer 40
Db 94 TACACCGCCAGCCCACTCAAGATATGACAGACCAACCGAGCAATATGGGCAACAAAGC 153
QY 41 TyrGlyThrTyrGlyGlnProThrAspValSerTyrThrGlnAlaGlnThrThrAlaThr 60
Db 154 TATGAACCTATGACAGCCCACTGATGCTATTAACCAAGGCTCAGACCTGCAACC 213
QY 61 TyrGlnGlnThrAlaTyrAlaThrSerTyrGlyGlnProProThrGlyTyrThrThrPro 80
Db 214 TATGGGCAACCGCTTATGACACTTCTTATGACAGCTTCCACTGTTTACTACTCCA 273
QY 81 ThrAlaProGlnAlaTyrSerGlnProValGlnGlyTyrGlyThrGlyAlaTyrAspThr 100
Db 274 ACTGCCCCAGGCAATACAGCCAGCGCTGTCAGAGGATATGACATGTGCTTATGATACC 333
QY 101 ThrThrAlaThrValThrThrThrGlnAlaSerTyrAlaAlaGlnSerAlaTyrGlyThr 120
Db 334 ACCACTGCTACAGTCAACCAACAGGCTCTCTTATGACAGTCACTGCTGATATGGCACT 393
QY 121 GlnProAlaTyrProAlaTyrGlyGlnGlnProAlaAlaThrAlaProThrArgProGln 140
Db 394 CAGCTGCTTATCAGCTTATGAGGAGAGAGCCAGAGCCAGCACTGACCTACAGAGCCGAG 453
QY 141 AspGlyAsnLysProThrGlnTyrSerGlnProGlnSerSerThrGlyLysTyrAsnGln 160
Db 454 GATGAAACAGCCCACTGAGACTTACATCACTTACAGCAGAGGGGTTTCAACAG 513
QY 161 ProSerLeuGlyTyrGlyGlnSerAsnTyrSerTyrProGlnValProGlySerTyrPro 180
Db 514 CCGAGCTGAGATATGACAGAGTAACTACAGTTATCCCAAGGTACTGAGAGACTACCCC 573
QY 181 MetGlnProValThrAlaProProSerTyrProProThrSerTyrSerSerThrGlnPro 200
Db 574 ATCAGCAGTCACTGACCTCATCTTACCTTCCCTTACAGCTATTCCTTACAGCCG 633
QY 201 ThrSerTyrAspGlnSerSerTyrSerGlnGlnAsnThrTyrGlyGlnProSerSerTyr 220
Db 634 ACTAGTTATGATCAGAGCACTTACTCTCAGCAGAAACCTATGGGCAACGAGCAAGTAT 693
QY 221 GlnGlnGlnSerSerTyrGlyGlnGlnSerSerTyrGlyGlnGlnProProThrSerTyr 240
Db 694 GACACAGAGTATGCTATGCTCAACAAACAGTATGGGAGAGAGCTCCCACTAGTTAC 753
QY 241 ProProGlnThrGlySerTyrSerGlnAlaProSerGlnTyrSerGlnGlnSerSer 260

```

[illegible][illegible]



```

Db      454  AATGAAACAGCCCACTGAGCTAGTCAACCTCACTGACAGAGGGGTTACACAG 513
      161  ProSerLeuGlyTyrGlyGlnSerAsnTyrSerTyrProGlnValProGlySerTyrPro 180
      514  CCCAGCTTAGATATGAGCAGAGTAACTGACGATTATCCCAAGGTAACCTGGAGTACTCC 573
Qy      181  MetGlnProValThrAlaProProSerTyrProProThrSerTyrSerSerThrGlnPro 200
      574  ATGACAGCACTACCCGACCTCCATCTCACTCAAGCTATCTCTTACACAGCCA 633
Qy      201  ThrSerTyrAspGlnSerSerTyrSerGlnGlnAsnThrTyrGlyGlnProSerSerTyr 220
      634  ACTAGTTATATACAGACAGCTTACTCTCAGACAGAACCTTATGGAGAAACGAGACCTTAT 693
Db      221  GlyGlnGlnSerSerTyrGlyGlnGlnSerSerTyrGlyGlnGlnProProThrSerTyr 240
      694  GACAGCAGAGTAGTATGCTATGCTCAACAAAGCAGCTATGGCAGCTGCTCCCACTAGTTAC 753
Qy      241  ProProGlnThrGlySerTyrSerGlnAlaProSerGlnTyrSerGlnGlnSerSer 260
      754  CCAACCCCAACTGATCTTCAAGCCAAAGCTCCAAAGTCAATAGCTTAAACAGAGCAGCAGC 813
Qy      261  TyrGlyGlnGlnSerSerPheArgGlnAspHisProSerSerMetGlyValTyrGlyGln 280
      814  TACGGCAGCAGAGA----- 828
Qy      281  GluSerGlyGlyPheSerGlyProGlyGlnAsnArgSerMetSerGlyProAspAsnArg 300
      828  ----- 828
Qy      301  GlyArgGlyArgGlyGlyPheAspArgGlyGlyMetSerArgGlyGlyArgGlyGly 320
      828  ----- 828
Qy      321  ArgGlyGlyMetGlySerAlaGlyGluArgGlyGlyPheAsnLysProGlyGlyProMet 340
      829  ----- 834
Db      341  AspGlnGlyProAspLeuAspLeuGlyProProValAspProAspGlyLysAspAspAsn 360
      835  GATGAAGGACCAAGATCTTGAATCTAGGCCCACTGATAGTCAAGTGAAGACTGCAAC 894
Qy      361  SerAlaIleTyrValGlnGlyLeuAsnAspSerValThrLeuAspLeuAlaAspPhe 380
      895  AGTGCATTTATGTACAAAGATTAAATGACATGTGACTAGATGATCTGTAAGACTTC 954
Qy      381  PheLysGlnCysGlyValValLysMetAsnLysArgThrGlyGlnProMetIleHisIle 400
      955  TTTAAGCAGTGTGGGTTGTTAGATBACAAAGAACTAGACAACTCATGATTCACACC 1014
Qy      401  TyrLeuAspLysGlnThrGlyLysProLysGlyAspAlaThrValSerTyrGlyLysPro 420
      1015  TACTGTGACAGGAACAAAGAAAGCCCAAGGTGATGCCACAGTCTCTTGAAGACTCA 1074
Db      421  ProThrAlaLysAlaLysValGluTTPheAspGlyLysAspPheGlnGlySerLysLeu 440
      1075  CCTACTGCTCCAAAGCTGCGTGAATGTTGATGGGAAAGATTTCAGAGGAGCAAACTT 1134
Qy      441  LysValSerLeuAlaAspLysLysProProMetAsnSerMetLysArgGlyGlyLysProPro 460
      1135  AAAGTCTCTTCTCGAAAGAGGCTCCAGTGAACGTAATGACAGGTGTATGACCACTCC 1194
Qy      461  ArgGlnGlyArgGlyMetProProProLeuArgGlyGlyProGlyGlyProGlyGlyPro 480
      1195  CGTAGGAGGAGAGGAGTGCACCACTCTGCGGAGGTCCAGAGGCCCAAGAAAGTCTT 1254
Qy      481  GlyGlyProMetGlyArgMetGlyGlyArgGlyGlyAspArgGlyGlyPheProProArg 500
      1255  GGGGAGACCCATGGGTCAATGGAGGCGCTGAGAGAGATAGAGAGGCTCCCTCCAAAGA 1314
Qy      501  GlyProArgGlySerArgGlyAsnProSerGlyGlyGlyAsnValGlnHisArgAlaGly 520
      501  |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```

Db      1315  GACCCCAAGGTTCCCGAAGGAACACCTCTGAGAGAGAAAGCTCCACAGCAAGCTGA 1374
Qy      521  AspTyrGlnCysProAsnProGlyCysGlyAsnGlnAsnPheAlaTyrPargThrGlyCys 540
      1375  GACAGCAGAGTGTCCCAATCCGGGTGTGGAACACAGAACTTCCCTGGAAGACAGAGAGC 1434
Qy      541  AsnGlnCysLysAlaProLysProGlnGlyPheLeuProProPheProProProGly 560
      1435  AACAGGTAAAGGCTCCAAAGCCTGAAGGCTTCTCCCGCACCTTCCACCCCGGGGT 1494
Qy      561  GlyAspArgGlyArgGlyGlyProGlyGlyMetArgGlyGlyArgGlyGlyLysMetAsp 580
      1495  GTGATCATGCGCAGAGTGGCCCTGTGTGATGTGGAGAGAAAGAGTGGCTCATGAT 1554
Db      581  ArgGlyGlyProGlyGlyMetPheArgGlyGlyArgGlyGlyAspArgGlyGlyPheArg 600
      1555  CATGTGTGTCCTCCGTGAATGTTCAAGAGGTGCTGTGTGAGAAACAGAAATGGCTTCCT 1614
Qy      601  GlyGlyArgGlyMetAspArgGlyGlyPheGlyGlyGlyArgArgGlyGlyProGlyGly 620
      1615  GTTGGCTG-GGCATGACCGAGGTGGCTTGTGTGAGAGAAAGACAAAGTGGCCCTGGGGGG 1673
Qy      621  ProProGlyProLeuMetGlnGlnMetGlyGlyArgArgGlyGlyArgGlyGlyProGly 640
      1674  CCCCCGGA-CTTTGATGTAACTGAATGGAGAGAAAGAGAGAGAGAGAGAGAGAGAGAGAG 1732
Db      641  LysMetAspLysGlyGlnHisArgGlnGluArgArgAspArgProTyr 656
      1733  AAAACGATTAAGCGCAGCAGCTGTCAAGAGCGCAGAGATCAGCCCTTAC 1780
Qy      656  ----- 1780
Db      1733  AAAACGATTAAGCGCAGCAGCTGTCAAGAGCGCAGAGATCAGCCCTTAC 1780
Qy      1733  ----- 1780

```

RESULT 6  
 US-08-437-027-18  
 ; Sequence 18, Application US/08437027  
 ; Patent No. 5670317  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Landanyi, Marc  
 ; TITLE OF INVENTION: A DIAGNOSTIC TEST FOR TEST FOR THE DESMOPLASTIC  
 ; TITLE OF INVENTION: SMALL ROUND CELL TUMOR  
 ; NUMBER OF SEQUENCES: 21  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Cooper & Dunham LLP  
 ; STREET: 1185 Avenue of the Americas  
 ; CITY: New York  
 ; STATE: New York  
 ; COUNTRY: U.S.A.  
 ; ZIP: 10036  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/437,027  
 ; FILING DATE:  
 ; CLASSIFICATION: 536  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: White, John P  
 ; REGISTRATION NUMBER: 28,678  
 ; REFERENCE/DOCKET NUMBER: 46416/JPW/CCA  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 212-278-0400  
 ; TELEFAX: 212-391-0525  
 ; INFORMATION FOR SEQ ID NO: 18:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 2413 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: cDNA to mRNA  
 ; US-08-437-027-18  
 Alignment..Scores:

Pred. No.: 1.35e-67 Length: 2412  
 Score: 1420.00 Matches: 265  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 39.09% Indels: 0  
 DB: 1 Gaps: 0

US-10-791-017a-2 (1-656) x US-08-437-027-18 (1-2412)

QY 1 MetAlaSerThrAspTyrSerThrTyrSerGlnAlaAlaAlaGlnGlnGlyTyrSerAla 20  
 DB 25 ATGGCGCTCCACGAGTACAGTACCTATAGCCAGCTCAGGCGACGAGGCTCAGAGTCT 84  
 QY 21 TTTTThraAlaGlnProThrGlnGlyTyrAlaGlnThrThrGlnAlaTyrGlnGlnSer 40  
 DB 85 TACACCGCCGACCGCTCAGAGTATGACAGACCCAGCCAGCATATGGGACACAAAGC 144  
 QY 41 TTTGlyThrTyrGlnProThrAspValSerTyrThrGlnAlaGlnThrThrAlaThr 60  
 DB 145 TATGGAACTATGACAGCCACGATGTACGCTATACCGAGCTCAGACCACTGCAACC 204  
 QY 61 TTTGlyGlnThrAlaTyrAlaThrSerTyrGlnGlnProProThrGlyTyrThrPro 80  
 DB 205 TATGGGACAGCCGCTATGCACTTCTATGACAGCTCCAGCTGTTTACTACTCCA 264  
 QY 81 ThrAlaProGlnAlaTyrSerGlnProValGlnGlnGlyTyrGlnGlnAlaTyrAspThr 100  
 DB 265 ACTGCCCCCGAGGATACAGCCAGCTGTCCAGGGATATGGCACTGGTCTTATGATACC 324  
 QY 101 ThrThraAlaThrValThrThrThrGlnAlaSerTyrAlaAlaGlnSerAlaTyrGlyThr 120  
 DB 325 ACCGCTGTAAGTACACCAACCAAGCCCTCCTATGAGTCACTGCTGATATGAGACT 384  
 QY 121 GlnProAlaTyrProAlaTyrGlyGlnGlnProAlaAlaThrAlaProThrAspGln 140  
 DB 385 CAGCTGCTTATCCAGCTATGAGGACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 444  
 QY 141 AspGlyAsnLysProThrGlnThrSerGlnProGlnSerSerThrGlyTyrAsnGln 160  
 DB 445 GATGAAACAAAGCCCACTAGACCTATGACCTTAACTTACGACAGGGGATTAACACAG 504  
 QY 161 ProSerLeuGlyTyrGlnGlnSerAsnTyrSerTyrProGlnValProGlySerTyrPro 180  
 DB 505 CCCAGCCATGAGTATGACAGAGTATGACAGTATATCCCAAGTACCTGAGAGTACCCC 564  
 QY 181 MetGlnProValThrAlaProProSerTyrProProThrSerTyrSerSerThrGlnPro 200  
 DB 565 ATGACGACAGTACAGTCACTCCATCCCTCACTCACTCACTCACTCACTCACTCACT 624  
 QY 201 ThrSerTyrAspGlnSerSerTyrSerGlnGlnAlaThrTyrGlnGlnProSerSerTyr 220  
 DB 625 ACTAGTATGATGACAGAGTATGCTTCAAGCAAAACCTTATGGGACAAACGAGAGCTAT 684  
 QY 221 GlyGlnGlnSerSerTyrGlnGlnGlnSerSerTyrGlnGlnGlnProProThrSerTyr 240  
 DB 685 GGAAGAGAGTATGCTATGCTCAACAAAGAGCTATGGGAGAGAGCTCCCACTAGTATAC 744  
 QY 241 ProProGlnThrGlySerTyrSerGlnAlaProSerGlnTyrSerGlnGlnSerSerSer 260  
 DB 745 CCACCCCAAACTGATCTTACAGCCAAAGCTCAAGTCAATATACCAACAGAGAGAGC 804  
 QY 261 TTTGlyGlnGlnSer 265  
 DB 805 TACGGGACAGAGAGT 819

# RESULT 7

US-09-919-039-322  
 Sequence 322 Application US/09919039  
 Patent No. 6727066  
 GENERAL INFORMATION:  
 APPLICANT: Kaber, Matthew R.  
 TITLE OF INVENTION: GENES EXPRESSED IN TREATED HUMAN C3A LIVER CELL CULTURES  
 FILE REFERENCE: PA-0035 US

CURRENT APPLICATION NUMBER: US/09/919,039  
 CURRENT FILING DATE: 2002-09-09  
 PRIOR APPLICATION NUMBER: 60/222,113  
 PRIOR FILING DATE: 2000-07-28  
 NUMBER OF SEQ ID NOS: 401  
 SOFTWARE: PERL Program  
 SEQ ID NO 322  
 LENGTH: 1939  
 TYPE: DNA  
 ORGANISM: Homo sapiens  
 FEATURE:  
 NAME/KEY: misc feature  
 OTHER INFORMATION: Incyte ID No. 6727066 478620.53  
 US-09-919-039-322

Alignment Scores:  
 Pred. No.: 7.51e-60 Length: 1939  
 Score: 1273.50 Matches: 301  
 Percent Similarity: 53.96% Conservative: 60  
 Best Local Similarity: 44.99% Mismatches: 153  
 Query Match: 35.05% Indels: 156  
 DB: 4 Gaps: 30

US-10-791-017a-2 (1-656) x US-09-919-039-322 (1-1939)

QY 1 MetAlaSerThrAspTyrSerThrTyrSerGlnAlaAlaAlaGlnGlnGlyTyrSerAla 20  
 DB 79 ATGGCGCTCCACGAGTATGACCTATAGCCAGCTCAGGCGACGAGGCTCAGAGTCT 126  
 QY 21 TTTTThraAlaGlnProThrGlnGlyTyrAlaGlnThrThrGlnAlaTyrGlnGlnSer 39  
 DB 127 TACCCACCAAGCCGAGGAGGCTATCCAGAGAGAGTACGCTTACGAGAGCAG 186  
 QY 40 SerTyrGlyThrTyrGlnGlnProThrAspValSerTyrThrGlnAlaGlnThrAla 59  
 DB 187 AGTTACAGTGTATGACAGTCCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 222  
 QY 60 ThrTyrGlyGlnThrAlaTyrAlaThrSerTyrGlyGlnProProThrGlyTyrThr 78  
 DB 223 GGGTATGGCCAGAGCAGCTATCTTCTTATGGCAGACCAAGCAACAGCTATGGA 279  
 QY 79 ThrProThrAlaProGlnAlaTyrSerGlnProValGlnGlyTyrGlyThrGlyAla 97  
 DB 280 ACTCAGTCAACTCCC-----CAGGGATATGCTGCACTGCGCGC 318  
 QY 98 TyrAspThrThrThrAlaThrValThrThrThrGlnAlaSerTyrAlaAlaGlnSerAla 117  
 DB 319 TATGGC-----AGTACGAGAGCTCCCAATGCTC 348  
 QY 118 TTTGlyThrGlnProAlaTyrProAlaTyrGlyGlnGlnProAlaAlaThrAlaProThr 137  
 DB 349 TACGGGACAGAGCTCTCTCTGCTATGGCAGACAGCAGCT----- 393  
 QY 138 ArgProGlnAspLysAsnLysProThrGlnThrSerGlnProGlnSerSerThrGly 157  
 DB 394 -----CCGAGCAGACCTCGGAAAGT 414  
 QY 158 TyrAsnGlnProSerLeuGlyTyrGlnGlnSerAsnTyrSerTyrProGlnValProGly 177  
 DB 415 TACGGTACAGTCT-----CAGAGCAGAGCTATGGGAGAGCCCAAG-----AGTGGG 462  
 QY 178 SerTyrProMetGlnProValThrAlaProProSerTyrProProThrSerTyrSerSer 197  
 DB 463 AGCTACAGCAGAGCT-----AGCTATGATGCA 492  
 QY 198 ThrGlnProThrSerTyrAspGlnSerSerTyrSerGlnGlnAlaThrTyrGlnPro 217  
 DB 493 -----CAGCAGCAAGCTATGACAGAGCAAAAGCTATATCCCTC 534  
 QY 218 SerSerTyrGlyGlnGlnSerSerTyrGlyGlnGlnSerSerTyrGlyGlnGlnProPro 237  
 DB 535 CAGGGCTATGACAGAGCAACAGTACCAACAGCAGCAGAGTGTGTGAGAGTGTGAGGT- 593

QY 238 ThSerTyProProGlnThrGlySerGlnAlaProSerGlnIrrSerGln 257  
 DB 594 GAGAGT-----GAGAGTACTATGCGCAATCAATCTCCATGAGTAGTGT 641  
 QY 258 SerSerSerTyGlnGlnSerSerPheArgGlnAphIAProSerSerMetGlyVal 277  
 DB 642 GGTGGAGGTGTGGTGGTGTATGCGAATCAAGACAGAGTGTGGAGGTGGAGCGGTGGC 701  
 QY 278 TyrGlyGlnGlnSerGlyGlyPheSerGlyProGlyGlnAphSerMetSerGlyPro 297  
 DB 702 TATGGACAGACGACCGTGA-----GGCGCGCGACGGGTGGCAGGTGT--- 746  
 QY 298 AspAphArgGlyArgGlyArgGlyGlyPheAphArg-----GlyGlyMetSerArgGly 315  
 DB 747 GGGGGCGCGCGCGCGCGGTGTGTGTACACCGCACGCGTGTGTGTGAACCCAGA 806  
 QY 316 GlyArgGlyGlyGlyArgGlyGlyMetGlySerAlaGly-----GluArgGlyGlyPhe 333  
 DB 807 GGTGTGGAGGTGTGGTGGTGGAGGTGGAGGTGGAGGTGGAGGTGGAGGTGGAGGTGGC 866  
 QY 334 AsnLysProGlyGlyProMetAspGlyGlyProAspLysAphGlyProProValAsp 353  
 DB 867 AATTAATTTGT 911  
 QY 354 ProAspGlnAphSerAspAphSerAlaIleTyrValGlnGlyLysAphSerValThr 373  
 DB 912 GAAACGATATATCAACCAACCAACCACTTTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 971  
 QY 374 LeuAspLysPheLysPhePheLysGlnCysGlyValValLysMetAsnLysArgThr 393  
 DB 972 ATTAGAGT 1031  
 QY 394 GlyLysProMetIleIleIleTyrLeuAspLysGlyGlyPheProLysPheValAspAla 413  
 DB 1032 GGACAGCCCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1091  
 QY 414 ThrValSerTyGlnAphProProThrAlaLysAlaValGlnTTPheAspGlyLys 433  
 DB 1092 ACGGTCTTTGT 1151  
 QY 434 AspPheGlnGlySerLysLysValSerLysLysValSerLysLysValSerLysLys 453  
 DB 1152 GAATTCGCCGAATCTATCAAGTCTCATTTGCTACTGCGCGGACGACTTTAAT--- 1208  
 QY 454 MetArgGlyLysLeuProProArgGlyGlyArgGlyMetProProLysArgGlyLys 473  
 DB 1209 ---CGGGGT 1250  
 QY 474 ProGlyGlyProGlyGlyProGlyGlyProMetGlyArgMetGlyGlyArgGlyLys 493  
 DB 1251 CCCATGGGCGGTGGAGGT 1298  
 QY 494 ArgGlyGlyPheProProArgGlyProArgGlySerArgLysAnProSerGlyGlyLys 513  
 DB 1299 CGAGAGGAGTATCCCACTGAGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1331  
 QY 514 AsnValGlnIleValArgAlaGlyAspTyrGlnCysProAsnProGlyGlyArgLysAn 533  
 DB 1332 GAGAGACAGACGACGAGT 1391  
 QY 534 PheAlaLysPheArgGlyGlyAsnGlnCysLysAlaProLysProGlyGlyLysPheLeuPro 553  
 DB 1392 TTCTCTTGGAGGAGT 1442  
 QY 554 ProProPheProProGlyGlyAspArgGlyArgGlyGlyProGlyGlyLysMetArg--- 572  
 DB 1443 ---CCAGAG-----GGGGGACCAAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1472  
 QY 573 GlyGlyArgGlyGlyLysMetAspArgGlyGlyProGlyGlyLysMetPheArgGly--- 590  
 DB 1473 GGGGGTAACTACGCGGAGT 1532  
 QY 591 ---GlyArgGlyGlyAspArgGlyGlyPheArgGlyGlyLys---GlyMetAspArgGly 608

DB 1533 CGGGGGCGCGCGCGGACCGT 1592  
 QY 609 GlyPheGlyGlyArgGlyGlyProGlyGlyProProGlyProLysMetGln 628  
 DB 1593 GCGTTT----- 1598  
 QY 629 MetGlyLysArgGlyGlyArgGlyGlyProGlyGlyMetAsp---LysGlyGlnIle 647  
 DB 1599 ---GGCGTGGCAAGT 1631  
 QY 648 ArgGlnGlnArgArgAspArgProTyr 656  
 DB 1632 AGACAGATGCGACGAGAGCGCGTAT 1658  
 RESULT 8  
 US-09-949-016-16785  
 ; Sequence 16785, Application US/09949016  
 ; Patent No. 6812339  
 ; GENERAL INFORMATION:  
 ; APPLICANT: VENTER, J. Craig et al.  
 ; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
 ; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
 ; FILE REFERENCE: C1001307  
 ; CURRENT APPLICATION NUMBER: US/09/949, 016  
 ; CURRENT FILING DATE: 2000-04-14  
 ; PRIOR APPLICATION NUMBER: 60/241,755  
 ; PRIOR FILING DATE: 2000-10-20  
 ; PRIOR APPLICATION NUMBER: 60/237,768  
 ; PRIOR FILING DATE: 2000-10-03  
 ; PRIOR APPLICATION NUMBER: 60/231,498  
 ; PRIOR FILING DATE: 2000-09-08  
 ; NUMBER OF SEQ ID NOS: 207012  
 ; SOFTWARE: FASTSEQ for Windows Version 4.0  
 ; SEQ ID NO 16785  
 ; LENGTH: 35784  
 ; TYPE: DNA  
 ; ORGANISM: Human  
 ; US-09-949-016-16785  
 Alignment Scores:  
 Pred. No.: 1,276-41 Length: 35784  
 Score: 958.00 Matches: 233  
 Percent Similarity: 29.75% Conservative: 5  
 Best Local Similarity: 29.12% Mismatches: 11  
 Query Match: 26.37% Indels: 554  
 DB: 4 Gaps: 6  
 US-10-791-017A-2 (1-656) x US-09-949-016-16785 (1-35784)  
 QY 402 LeuAspLysGlnThrGlyLysProLysGlyAspAlaThrValSerTyGlnAphProPro 421  
 DB 31405 TTGACAGAACTTACAGACCACTTAATGATCGGAGGTGTGTGTGTGTGTGTGTGTGT 31460  
 QY 422 ThrAlaLysAlaValGlnTTP--- 429  
 DB 31461 ---GTGGAGT 429  
 QY 430 ---PheAspGlyLysAspPheGlnGlySerLysLysValSerLysLysValSerLysLys 447  
 DB 31506 TTGTCTTGTTCACGAGGAAATTTTCAAGGACCAAACTTAAGTCTCCCTTGTGTGTGA 31565  
 QY 447 LysProProMetAsnSerMetArgGlyGlyLysLeuProProArgGlyGlyArgGlyMetPr 467  
 DB 31566 GAAGCCCTCAATGAACAGTATGCGGGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 31625  
 QY 467 oProProLysArgGly--- 472  
 DB 31626 ACCACCACTTCGAGAGGTATCTTTTGTGACTCTATGTTCATTAAAGTTTTCAGT 31685  
 QY 472 --- 472  
 DB 31686 ACACTTCATACCTTGAGAACTTGATTTAAGTGAAGAAATTAATTAATTTGTGTGTAG 31745



CURRENT APPLICATION NUMBER: US/09/949,016  
CURRENT FILING DATE: 2000-04-14  
PRIOR APPLICATION NUMBER: 60/241,755  
PRIOR FILING DATE: 2000-10-20  
PRIOR APPLICATION NUMBER: 60/237,768  
PRIOR FILING DATE: 2000-10-03  
PRIOR APPLICATION NUMBER: 60/231,498  
PRIOR FILING DATE: 2000-09-08  
NUMBER OF SEQ ID NOS: 207012  
SOFTWARE: FASTSEQ for Windows Version 4.0  
SEQ ID NO 16786  
LENGTH: 35784  
TYPE: DNA  
ORGANISM: Human  
US-09-949-016-16786

Alignment Scores:  
Pred. No.: 1,27e-41 Length: 35784  
Score: 958.00 Matches: 233  
Percent Similarity: 29.75% Conservative: 5  
Best Local Similarity: 29.12% Mismatches: 11  
Query Match: 26.37% Indels: 554  
Gaps: 6

US-10-791-017a-2 (1-656) x US-09-949-016-16786 (1-35784)

QY 402 Leuaplysglunthrglylybprolybgllyasplalathvalserlyrgluappropro 421  
Db 31405 TTAGAGAAAGATTACAGGACAGACCTAATGACATCTGGAGGTG-TCAATTGATGATGAT- 31460  
QY 422 Thralalysalalalvalglu-TTP----- 429  
Db 31461 -----GTGAGATTGTGTAGACAGAGGATACAGGGAGTAAATGATGTTCTG 31505  
QY 430 -----Pheaspolylyasapppheglnglyserlyleuylvalserleu1aargly 447  
Db 31506 TTGCTCTGTTCCAGGAGAAAGATTTCAGAGGAGCAACTTAATCTCTCCCTGCTCGAA 31565  
QY 447 blybproprometanbsermetarglylyleuoproproarglylyarglymetpr 467  
Db 31566 GAAGCCTCCATGAAAGATATGGGGGTGTCTGCCACCCCGTAGGGCAGAGGCATGCC 31625  
QY 467 oProProleuargly----- 472  
Db 31626 ACCACCACTCCGTGAGGTACTTTTCTGAGCTCCTATGTCATTAAAGTTTTCAGT 31685  
QY 472 ----- 472  
Db 31686 ACACTTCATACCTTGAGAACTTGATTAATTAGAGTAGAAGAAATATATAAATTGTGTAG 31745  
QY 472 ----- 472  
Db 31746 AGTCAATACTAGACTATCGAGAGCTAACATGATGTTTGTGGGAATTAAGAGAGAGA 31805  
QY 472 ----- 472  
Db 31806 AGAAGATGGAGGCTGAGAGCCACTGCTGCTCACTCCAGACTGCCATTATTCAGCT 31865  
QY 472 ----- 472  
Db 31866 TTGGTGTGTCTGTATAGACATGCCCTATTCCTTAAGAAATTGGGAGTTCCAGCCAGGTG 31925  
QY 472 ----- 472  
Db 31926 CAGTGCTACGCTGTATATCCAGCACTTGGAGAGGTAGAGTGGGGATCACTGAG 31985  
QY 472 ----- 472  
Db 31986 GTCAGAGTTTAAGACAGCCTGCGCAACGTGTGAAATCCTCTCTACTAATAAATACA 32045  
QY 472 ----- 472  
Db 32046 AATAATTAGCCGGGTATGTGTGTGATGCTGTATAATCCAGCTACTCGGAGGCTGAGGCA 32105

QY 472 ----- 472  
Db 32106 GAGAAATTGTTGAATCTGGGGGGGTGAGAGTTGCACTGATGAGCAAAAGATCGTCCACTGCA 32165  
QY 472 ----- 472  
Db 32166 CTCAGCCTGGCAACAGTGTGAGACTCCCTCTCAAAAAAAAAAAAAAAAAAATTGTGGG 32225  
QY 472 ----- 472  
Db 32226 AGCTGTGTTTCTGTAGAGACGTGAACACGCTCTCTCACAGGAAAGGGGCTGATGGCT 32285  
QY 472 ----- 472  
Db 32286 GAGCCACACGAAAAACGGGACAGGTGATGGGAAATGACAGCATAGTATCTGTGGTT 32345  
QY 472 ----- 472  
Db 32346 TACTTAGTATTTTATTTCTTATAGCAAAATTTGGTCTACAGAAATGATTTGCTGTT 32405  
QY 473 ----- 473  
Db 32406 TCTTGTGTTCTTGTGTAGTCCAGGAGGCCACAGAGGTCTGGGGGACCATGGGTG 32465  
QY 486 gMetglylyarglylylyasparglylylypheproproarglylyproarglylyserar 506  
Db 32466 CATGGGAGCCGTGAGAGATAGAGAGGCTTCCCTCCAAAGAGACCCCGGGGTTCCCG 32525  
QY 506 gglYasProSerGlylylylyasvalglnhlsarYalaglYAspThglnCySProAs 526  
Db 32526 AGGAAACCCCTGTGAGAGAGAAAGCTCCAGCACCGAGCTGAGAGACTGGAGTGTCCAA 32585  
QY 526 nProGly----- 528  
Db 32586 TCCGTA-TGTACTGTCTTGCAAAATTGATACCTACGAGTGAAGCAACCCCTCCCTCAC 32644  
QY 528 ----- 528  
Db 32645 CCCATCCCACTAGAGTGAATGCTCTGTCTAGAGAAAGAAATGATGACCTGATGG 32704  
QY 528 ----- 528  
Db 32705 CTGTTAGGACACTAGTACGCAATTCATCTGACGCTTACAGACCTTCTGAAGATTGATT 32764  
QY 528 ----- 528  
Db 32765 TGACTGTCTGTGGGTGCAATGCTGCTGAGGCTGTGCTTAAGCATGGGTATACATA 32824  
QY 528 ----- 528  
Db 32825 GATCTCTGTATGATGATGTGATACCTGTTACACACACACTTTCCTGTTTATCTTCTCT 32884  
QY 529 ----- 529  
Db 32885 TACCTAATTGTGATTTCTGTGTATGATGAATGATGATGAGGGGTTGTGAAACAGAA 32944  
QY 533 nPhealATPARGThrglylyCyshenglnCylysaliaProlybProgllylyPheleupr 553  
Db 32945 CTTCGCTGAGAAACAGAGTGCACAGTGTAAAGCCCCCAAAAGCTGAAAGCTTCTCTCC 33004  
QY 553 oProProPheProPro----- 559  
Db 33005 GCCACCTTTCGCCCCCGGTAGTGCAGATTTCATGAGTGTCCCTCAGCTTCTGTGT 33064  
QY 559 ----- 559  
Db 33065 GCTAAACCTCTTTCTTATTTGTGGGCTGTGTAATGACAGTTGCCCTCTGCTTAACAC 33124  
QY 559 ----- 559  
Db 33125 TTGAGTTGTGTGTCTCTCATTTCTAATTGTGAGCCCAATGCCGAGATTGAGTGAAGTG 33184

```

US-10-791-017A-2 (1-656) x US-09-513-999C-1657 (1-456)
QY      1 MetAlaSerThrAspTyrSerThrTyrSerGlnAlaAlaAlaGlnGlnIYTySerAla 20
Db      56 ATGGGCTCACGGATTATACAGTACTATAGCCAAAGCTGACGGCA -CAGGGCTACAGTCT 114
QY      21 TyrThrAlaGlnProThrGlnGlnIYTyAlaGlnThrThrGlnAlaIYTyGlnGlnSer 40
Db      115 TACACCGGCCACGGCCATCAAGGATATGCACAGACCACCCAGGCAATATGGCAACAAAGC 174
QY      41 TyrGlnThrTyrGlnIYTyGlnProThrAspValSerTyrThrGlnAlaGlnThrThrAlaThr 60
Db      175 TATGGAACTTATGACACGCCCACTATGTCACTTATCCAGGCTCAGACCACTGCAACC 234
QY      61 TyrGlnGlnThrAlaIYTyAlaThrSerTyrGlnGlnProProThrGlnIYTyThrThrPro 80
Db      225 TATGGGACAGACCGGCTATGCAACTTCTTATGGACAGCCCTCCACATGGTTATACTCTCCA 294
QY      81 ThrAlaProGlnAlaIYTySerGlnProValGlnGlnIYTyGlnIYTyAlaIYTyAspThr 100
Db      295 ACTGGCCCCCAGGCAATACAGCCAGCCCTGTCCAGGGGTAATGGACATGGTCTTATATGATACC 354
QY      101 ThrThrAlaThrValThrThrThrGlnAlaSerTyrAlaAlaGlnSerAlaIYTyGlnIYThr 120
Db      355 ACCACTGCTTACAGTACCAACCAACCCAGGCTCTCTTATGCAAGCTCAAGTGTGCAATATGGCACT 414
QY      121 GlnProAlaIYTyProAlaIYTyGlnIYGlnIYProAlaAla 133
Db      415 CAGCGTCTTATCCAGCGCTATATGGGACGACGACGACGACC 453

RESULT 11
US-09-621-976-13361
; Sequence 13361, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Giordano, S.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 13361
; LENGTH: 411
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-621-976-13361

Alignment Scores:
Pred. No.:      5,55e-25      Length:      411
Score:          609.00      Matches:      116
Percent Similarity: 99.15%      Conservative: 0
Best Local Similarity: 99.15%      Mismatches: 0
Query Match:      16.76%      Indels:      1
DB:               Gaps:      0

US-10-791-017A-2 (1-656) x US-09-621-976-13361 (1-411)
QY      338 G1YProMetAaPgiGlnIYProAspLeuAspLeuG1YProProValAspProAspG1uaAp 357
Db      62  GAACCCATGATGTAAGAGACCAAGATCTTGATCTAGGCCCACTGTAGATCCAGATGAAAGAC 121
QY      358 SerAspAsnSerAlaIleTyrValGlnIYLeuAsnAspSerValThrIleuAspLeu 377
Db      122  TCTGACCAACAATGCAATTATGTAACAGGATTAATATGACAGGTATCTAGATGATCTG 181
QY      378 AlaAspPhePheIYTyGlnCYGlnIYValIYValIYbMetAsnIYbArgThrGlnIYProMet 397
Db      182  GCAAGCTTCTTTAAGCAGTGTGGGCTTTTAAAGATGAACAGAGAACTGGGCAACCCATG 241

```



```

Qy 398 ILEHSIIETYLELEASPPYSGITRNGLYLVSProLYSGIYASP-ALATHrValseTy 417
Db 242 ATCCACATCTACTGTGACAGAAACAGAAAGCCAAAGGCGATTGCCACATGTCTCA 301
Qy 417 TGLuAPProPOThrAlaLYaLaVaIGutTPheASPGLYLYaSPheGInGI 437
Db 302 TGAAGACCAACCACTGCCAGAGCGTCCGTGGAATGTTGAATGGAAAGATTTTCAAG 361
Qy 437 YSErLYaLeuLYaValSERLeuAlaRGLYsLYaSPProMetLeuSer 453
Db 362 GAGCAAACTTAAAGTCTCCCTGTGCGAAGAAAGCTCCCAATGACAGT 410

RESULT 12
US-09-949-016-66382/C
Sequence 66382, Application US/09949016
Patent No. 6812339
GENERAL INFORMATION:
APPLICANT: VENTER, J. Craig et al.
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
FILE REFERENCE: CL001307
CURRENT FILING DATE: 2000-04-14
PRIOR APPLICATION NUMBER: 60/241,755
PRIOR FILING DATE: 2000-10-20
PRIOR APPLICATION NUMBER: 60/237,768
PRIOR FILING DATE: 2000-10-03
PRIOR APPLICATION NUMBER: 60/231,498
PRIOR FILING DATE: 2000-09-08
NUMBER OF SEQ ID NOS: 207012
SOFTWARE: PASTSEQ for Windows Version 4.0
SEQ ID NO 66382
LENGTH: 601
TYPE: DNA
ORGANISM: Human
US-09-949-016-66382

Alignment Scores:
Pred. NO.: 1,666-22 Length: 601
Score: 566.00 Matches: 107
Percent Similarity: 94.69% Conservaive: 0
Best Local Similarity: 94.69% Mismatches: 6
Query Match: 15.58% Indels: 0
DB: 4 Gaps: 0

US-10-791-017A-2 (1-656) x US-09-949-016-66382 (1-601)
Qy 1 MetaIaSerThrAspTYrSerThrTYrSerGlnAlaAlaGlnGlnGlyTYrSerAla 20
Db 341 ATGGATCCATGATTAACGTACTATACCAAGCGAGGAGAGAGGGCTACATGCT 282
Qy 21 TYrThraIaGlnProThrGlnGlyTYrAlaGlnThrThrGlnAlaTYrGlnGlnSer 40
Db 281 TACAACACCCAGCCCACTCAAGGATATGCAACAGACCAACCGACATATGGACAAAGC 222
Qy 41 TYrGlyThrTYrGlyGlnProThrAspValSerTYrThrGlnAlaGlnThrThraIaThr 60
Db 221 TATGGAACCTATGAGCAGCCCATGTGATGTGAGTATACCAAGGCTCAGACCACTGCAATC 162
Qy 61 TYrGlyGlnThraIaTYrAlaThrSerTYrGlyGlnProProThrGlyTYrThrPro 80
Db 161 TATGGCAGAGCCGCTATGCAACTTCTTATGAGCAGCCTCCCACTGTATATACTCTCCA 102
Qy 81 ThraIaProGlnAlaTYrSerGlnProValGlnGlyTYrGlyThrGlyAlaTYrAspThr 100
Db 101 ACTGCCCCCAGGACATACAGCCAGCCTGTCCAGGGGATATGGCACTGTGTCTTATGATACC 42
Qy 101 ThrThraIaThrValThrThrThrGlnAlaSerTYrAla 113
Db 41 ACCACTGTACAGTACCAACCAACCGGCTCTTATGCA 3

```

```

: Sequence 82, Application US/09220132
: Patent No. 6506607
: GENERAL INFORMATION:
: APPLICANT: Shyan, Andrew W.
: TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE IDENTIFICATION AND ASSESSMENT
: FILE OF INVENTION: OF PROSTATE CANCER THERAPIES AND THE DIAGNOSIS OF PROSTATE CANCER
: FILE REFERENCE: 07334-074001
: CURRENT APPLICATION NUMBER: US/09/220,132
: PRIOR FILING DATE: 1998-12-23
: PRIOR APPLICATION NUMBER: US 60/079,303
: PRIOR FILING DATE: 1998-03-25
: PRIOR APPLICATION NUMBER: US 60/068,821
: PRIOR FILING DATE: 1997-12-24
: NUMBER OF SEQ ID NOS: 191
: SOFTWARE: FASTSEQ for Windows Version 4.0
: SEQ ID NO 82
: LENGTH: 1682
: TYPE: DNA
: ORGANISM: Homo sapiens
US-09-220-132-82

Alignment Scores:
Pred. No.: 9,45e-18 Length: 1682
Score: 486.00 Matches: 176
Percent Similarity: 41.07% Conservative: 61
Best Local Similarity: 30.50% Mismatches: 182
Query Match: 13.38% Indels: 160
DB: 4 Gaps: 29

US-10-791-017A-2 (1-656) x US-09-220-132-82 (1-1682)
QY 1 MetAlaserThraspTySerThyrSerGlnAlaAlaGlnGlnGlyTyrSerAla 20
DB 79 ATGGCGCTCAAGGATTATAC-----CAACAAACAAACCAAGCTATGAGGCC 126
QY 21 TyrThrAlaGlnProThrGlnGlyTyrAla---GlnThrThrglnAlaTyrGlyGlnGln 39
DB 127 TACCCCAACCCACCCGGGGAGGCTATTCACAGAGAGCAGTCAGCCCTACGACAGCAG 186
QY 40 SerTyrGlyThrTyrGlyGlnProThrAspValSerTyrThrglnAlaGlnThrAla 59
DB 187 AGTTCAGTGCGTTATAGCCAGTCACGAC-----ACTTCA 222
QY 60 ThrTyrGlyGlnThrAlaTyrAlaThrSerTyrGlyGlnProPro---ThrglyTyrThr 78
DB 223 GGCCTATGGCCAGACGAGCTAT--TCTTCTATGGCCAGACCAAGAACAGCGCTATGGA 279
QY 79 ThrProThrAlaProGlnAlaTyrSerGlnProValGlnGlyTyrGly---ThrglyAla 97
DB 280 ACTCAGTCAACTCC-----CAGGAAATATGGCTGACATGGCGGC 318
QY 98 TyrAspThrThrThrAlaThrValThrThrThrglnAlaSerTyrAlaAlaGlnSerAla 117
DB 319 TATGGC-----AGTACGACAGCTCCCAATCGCT 348
QY 118 TyrGlyThrglnProAlaTyrProAlaTyrGlyGlnGlnProAlaAlaThrAlaProThr 137
DB 349 TACGGGACAGCACTCTTACCTGCTATGGCCAGACGACCT----- 393
QY 138 ArgProGlnAspGlyAlaAspProThrGlnThrSerGlnProGlnSerSerThrglyGly 157
DB 394 -----CCGACGACACCTCGGAGAGT 414
QY 158 TyrAsnGlnProSerLeuGlyTyrGlyGlnSerAsnTyrSerTyrProGlnValProGly 177
DB 415 TACGGTAGCAATCT-----CAGACGACAGCTATATGGACAGCCCAAG--AGTGGG 462
QY 178 SerTyrPrometGlnProValThrAlaProProSerTyrProProThrSerTyrSerSer 197
DB 463 AGCTACAGCCAGCAGCT-----AGCTATGATGGA 492
QY 198 ThrGlnProThrSerTyrAspGlnSerTyrSerGlnGlnAlaThrTyrGlyGlnPro 217

```

[illegible]

```

US-08-343-443B-106
; Sequence 106, Application US/08343443B
; Patent No. 5968734
; GENERAL INFORMATION:
; APPLICANT: Aurias, Alain
; APPLICANT: Delattre, Olivier
; APPLICANT: Desmaze, Chantal
; APPLICANT: Melot, Thomas
; APPLICANT: Peter, Martine
; APPLICANT: Ploougasel, Beatrice
; APPLICANT: Thomas, Gilles
; APPLICANT: Zuchan, Jessica
; TITLE OF INVENTION: NUCLEIC ACID CORRESPONDING TO A GENE OF
; TITLE OF INVENTION: CHROMOSOME 22 INVOLVED IN RECURRENT CHROMOSOMAL
; TITLE OF INVENTION: TRANSLATIONS ASSOCIATED WITH THE DEVELOPMENT OF CANCEROUS
; TITLE OF INVENTION: TUMORS, AND NUCLEIC ACIDS OF FUSION RESULTING FROM SAID
; NUMBER OF SEQUENCES: 129
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Weiser & Associates
; STREET: 230 South Fifteenth Street
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19102
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: AEDIT 1.0 DOS text editor
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/343,443B
; FILING DATE: 18-NOV-1994
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/FR93/00494
; FILING DATE: 19-MAY-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: FR 92/06123
; FILING DATE: 20-MAY-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Weiser, Gerard J.
; REGISTRATION NUMBER: 19,763
; REFERENCE/DOCKET NUMBER: 989,6121P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-875-8383
; TELEFAX: 215-875-8394
; INFORMATION FOR SEQ ID NO: 106:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 954 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..885
US-08-343-443B-106

Alignment Scores:
Pred. No.: 6,65e-18 Length: 954
Score: 484.00 Matches: 89
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 13.32% Indels: 0
DB: 2 Gaps: 0

US-10-791-017A-2 (1-656) x US-08-343-443B-106 (1-954)
07 237 ProThSerTyrrProGlnThrGlySerTyrSerGlnAlaProSerGlnTyrSerGln 256
DB 1 CCACATGATTAACCCACACCCAACTGATCTTACAGCCAAAGCTCCAAAGTCAATATAGCCAA 60

```

```

US-08-343-443B-106
; Sequence 106, Application US/08343443B
; Patent No. 5968734
; GENERAL INFORMATION:
; APPLICANT: Aurias, Alain
; APPLICANT: Delattre, Olivier
; APPLICANT: Desmaze, Chantal
; APPLICANT: Melot, Thomas
; APPLICANT: Peter, Martine
; APPLICANT: Plooungastel, Beatrice
; APPLICANT: Thomas, Gilles
; APPLICANT: Zuchman, Jessica
; TITLE OF INVENTION: NUCLEIC ACID CORRESPONDING TO A GENE OF
; TITLE OF INVENTION: CHROMOSOME 22 INVOLVED IN RECURRENT CHROMOSOMAL
; TITLE OF INVENTION: TRANSLATIONS ASSOCIATED WITH THE DEVELOPMENT OF CANCEROUS
; TITLE OF INVENTION: TUMORS, AND NUCLEIC ACIDS OF FUSION RESULTING FROM SAID
; NUMBER OF SEQUENCES: 129
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Weiser & Associates
; STREET: 230 South Fifteenth Street
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19102
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: AEDIT 1.0 DOS text editor
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/343,443B
; FILING DATE: 18-NOV-1994
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/FR93/00494
; FILING DATE: 19-MAY-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: FR 92/06123
; FILING DATE: 20-MAY-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Weiser, Gerard J.
; REGISTRATION NUMBER: 19,763
; REFERENCE/DOCKET NUMBER: 989.6121P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-875-8383
; TELEFAX: 215-875-8394
; INFORMATION FOR SEQ ID NO: 106:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 954 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..885
US-08-343-443B-106

Alignment Scores:
Pred. No.: 6,65e-18 Length: 954
Score: 484.00 Matches: 89
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 13.32% Indels: 0
DB: 2 Gaps: 0

US-10-791-017A-2 (1-656) x US-08-343-443B-106 (1-954)
07 237 ProThSerTyrrProGlnInhrGlySerTyrSerGlnAlaProSerGlnTyrSerGln 256
DB 1 CCACATGATTAACCCACACCCAACTGATCTTACAGCCAAAGCTCCAAAGTCAATATAGCCAA 60

```

QY 257 GlnSerSerSerTyrGlyGlnGlnSerSerPheArgGlnAspHisProSerSerMetGly 276  
DB 61 CAGAGCAGCAGCTAACGGGCGAGCAGAGTTTCATTCGACAGACCCAGCTAGCATGGGT 120  
QY 277 ValTyrGlyGlnGlnSerSerGlyPheSerGlyProGlyGlnAsnArgSerMetSerGly 296  
DB 121 GTTTATGGGCGAGAGTCTGAGAGTTTCCGACCAAGAGAACCCGAGCATGATGGC 180  
QY 297 ProAspAsnArgGlyArgGlyArgGlyPheAspArgGlyGlyMetSerArgGly 316  
DB 181 CCGATTAACCGGGCGAGGGGAGAGGGGGAATTGATGTGGAGGCATGACACAGAGTGGG 240  
QY 317 ArgGlyGlyGlyArgGlyGlyMetGly 325  
DB 241 CCGGAGAGAGACCGCGGTGGAATGGGA 267  
RESULT 15  
US-09-919-039-323  
/ Sequence 323, Application US/09919039  
/ Patent No. 6727066  
/ GENERAL INFORMATION:  
/ APPLICANT: Kaseer, Matthew R.  
/ TITLE OF INVENTION: GENES EXPRESSED IN TREATED HUMAN C3A LIVER CELL CULTURES  
/ FILE REFERENCE: PA-0035 US  
/ CURRENT APPLICATION NUMBER: US/09/919, 039  
/ PRIOR FILING DATE: 2002-09-09  
/ PRIOR APPLICATION NUMBER: 60/222,113  
/ NUMBER OF SEQ ID NOS: 401  
/ SOFTWARE: PERL Program  
/ SEQ ID NO 323  
/ LENGTH: 1684  
/ TYPE: DNA  
/ ORGANISM: Homo sapiens  
/ FEATURE:  
/ NAME/KEY: misc feature  
/ OTHER INFORMATION: Incyte ID No. 6727066 1813444CB1  
US-09-919-039-323  
Alignment Scores:  
Pred. No.: 1,21e-17 Length: 1684  
Score: 484.00 Matches: 177  
Percent Similarity: 40.68% Conservat: 61  
Best Local Similarity: 30.26% Mismatches: 173  
Query Match: 13.32% Indels: 176  
DB: 4 Gaps: 30  
US-10-791-017a-2 (1-656) x US-09-919-039-323 (1-1684)  
QY 1 MetAlaSerThrAspTyrSerThrTyrSerGlnAlaAlaGlnGlnGlyTyrSerAla 20  
DB 66 ATGGCCTCAACGATTTACC-----CAACAGCAACCCAAAGCTATGGGGC 113  
QY 21 TyrThrAlaGlnProThrGlnGlyTyrAla----GlnThrThrGlnAlaTyrGlyGln 39  
DB 114 TACCCACCCAGCCCGGCGAGGCTATTCACAGCAGAGCTCCCTACGAGCAGCAG 173  
QY 40 SerTyrGlyThrTyrGlyGlnProThrAspValSerTyrThrGlnAlaGlnThrAla 59  
DB 174 AGTTACAGTGTTTACCGCTCCAGCAGC-----ACTTCA 209  
QY 60 ThrTyrGlyGlnThrAlaTyrAlaThrSerTyrGlyGlnProPro---ThrGlyTyrThr 78  
DB 210 GGCATATGCCAGACACTAT---TCTTATAGCCAGAGCCAGAACACAGGCTATGGA 266  
QY 79 ThrProThrAlaProGlnAlaTyrSerGlnProValGlnGlnGlyTyrGly---ThrGlyAla 97  
DB 267 ACTCAGCACTCCC-----CAGGATATGGCTCGACTGGGGC 305  
QY 98 TyrAspThrThrThrAlaThrValThrThrThrGlnAlaSerTyrAlaAlaGlnSerAla 117  
DB 306 TATGGC-----AGTACCGAGAGCTCCCATCTCT 335

QY 118 TyrGlyThrGlnProAlaTyrProAlaTyrGlyGlnGlnProAlaAlaThrAlaProThr 137  
DB 336 TACGGCAGACAGTCCCTCATCTGCTATGGCCAGACAGCAGCT----- 380  
QY 138 ArgProGlnAspGlyAsnLysProThrGlnThrSerGlnProGlnSerSerThrGly 157  
DB 381 -----CCAGCAGACACTGGGAGT 401  
QY 158 TyrAsnGlnProSerLeuGlyTyrGlyGlnSerAsnTyrSerTyrProGlnValProGly 177  
DB 402 TAGGATAGCAGTCT-----CAGACAGCAGCTATGGCAGCCCAAG---AGTGG 449  
QY 178 SerTyrProMetGlnProValThrAlaProProSerTyrProProThrSerTyrSer 197  
DB 450 AGCTACAGCCAGCAGCT-----ACCTATGGTGA 479  
QY 198 ThrGlnProThrSerTyrAspGlnSerSerTyrSerGlnGlnAsnThrTyrGlyGlnPro 217  
DB 480 -----CAGCAGCAAGCTATGACAGCAGCAAGCTAATATCCCT 521  
QY 218 SerSerTyrGlyGlnGlnSerSerTyrGlyGlnGlnSer----- 230  
DB 522 CAGGCTATGACAGCAGAACCAAGTACACAGCAGAGTGTGTGGAGGTGAGTGA 581  
QY 231 -----SerTyrGlyGlnGlnProPro-----ThrSerTyrProProGlnThrGly 245  
DB 582 GGTGAGATACATATGATCCCAAGATCAATCTCCATGATGTGTGTGGCAGTGTGGC 641  
QY 246 SerTyrSerGlnAlaProSerGlnTyrSerGlnGlnSerSerSerTyrGlyGlnGlnSer 265  
DB 642 GGTATGGCANTACAGACAGAGTGTGAGTGGCGCGGTGCTATGACAGCAGAC 701  
QY 266 SerPheArgGlnAspHisProSerSerMetGlyValTyrGlyGlnGlnSerGlyPhe 285  
DB 702 CGTGAAGCGCGGCGAGCGGTGCGCATGTGGC-----GGCGCGGGCGC 746  
QY 286 SerGlyProGlyGlnAsnArgSerMetSerGlyProAspAsnArgGlyArgGly 305  
DB 747 GCGGTGTGTGTATCAACCGCAGCAGTGTGTGTATGAAACCCAGAGCTCTGCA----- 800  
QY 306 GlyPheAspArgGlyGlyMetSerArgGlyArgGlyGlyGlyValArgGlyGlyMetGly 325  
DB 801 -----GGTGGCGG---GGAGCAGAGGTGCAATGGC 830  
QY 326 SerAlaGlyGlnArgGlyGlyPheAsnLysProGlyGlyProMetAspGlu----- 342  
DB 831 GGAAGT---GACCGTGTGCTTCAATAATTGTGTGTCAAGAAAGATATCTT 887  
QY 343 -----GlyProAspLeuAspLeuGlyProProValAspProAspGlyAspSerAspAsn 360  
DB 888 CATACATCACACACTGTAAGACAGATGTGCTTTCCAGCTGATCAACTGACAGATG 947  
QY 361 SerAlaIleTyrValGlnGlyLeuAsnAspSerValThrLeuAspLeuAlaAspPhe 380  
DB 948 GCGAGTCACTATGCTTCTTCTTGGGACACTGTCCAGCTGGAGCTGGAAGCTGG 1067  
QY 381 Phe-LysGlnGlyValValAlaLysMetAsnLysAspGlnGlyGlnProMetIleHis 400  
DB 1008 TATGAGAGCTGCAAGAGAGCTCTTCTTCAGATGAAGAAATGGGGTACTATGTTTCACT 1067  
QY 400 eTyrLeuAspLeuGlnThrGlyLysProLysGlyAspAlaThrValSerTyrGlnAspPr 420  
DB 1068 C---CTGGAATGTAMAGAAAGAAATCTTC--ACCACACTT-----GACCC 1114  
QY 420 oProThrAlaValAlaValGlnThrPheAspGlyLysAspPheGlnGlySerLysLe 440  
DB 1115 T-----GCTTCTGTGCTGTGCTGACTGAGAGAGACCAACCAAGCA----- 1157  
QY 440 LysValSerLeuAlaArgLysLysProProMetAsnSerMetArgGlyLysLeuProPr 460  
DB 1158 -GAGGTCAACAGACCTCCACAGAGCCTCA--CTCTCCAGATTCCATTCAGAGCTCCT-- 1213

Oy	460	oargluclyarlgylmetpropprobleuarlglyclprcglgylprcglgyl--	479
Dd	1214	-----ggctcagaggagaaaggagaag	1236
Oy	480	-----ProglyProme	484
Dd	1237	ACCAAGGAGAACCAGGAAGGAACAAGATGTCAATCCCAAGCCCCGGCTGGAAAGC	1296
Oy	484	tglyArMeclyclYarTgLyIYaP-	493
Dd	1297	AGCGCATGAAGAGAAACAGAGAGATGAAGAAAGTGACACAGCTAAGTAAGAGA	1356
Oy	494	-----ArgglyYlPhneProProrTgLyIProArTgLySerArTgLyYaen--	508
Dd	1357	ATGAACGGCTCAAGCAAGCAAAATCGAGCCCT-----GACCAGGGAAGTAAGCGCGACTC	1410
Oy	509	-----ProserglyYl------GlyanValGIInHear	518
Dd	1411	GCCGAGCTCTGATTGCACCGCAATGTGAATTCGACCAAGCATGAACAATTGGAGACATCA	1470
Oy	518	galaglYasrTrp	522
Dd	1471	gtccccacattgg	1483

Search completed: February 21, 2005, 04:22:17  
Job time : 451.042 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus.p2n model

Run on: February 21, 2005, 02:07:46 ; Search time 3629.78 Seconds  
(without alignments)  
1068.188 Million cell updates/sec

Title: US-10-791-017A-2

Perfect score: 3633  
Sequence: 1 MASTDSTYSQAAAGQGYSA.....GGGKNDKGRHRRDRDPY 656

Scoring table:  
BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 5384158 seqs, 2955248155 residues

Total number of hits satisfying chosen parameters: 10768316

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODE=frame+ p2n.model -DEV=xlp  
-O=/cgn2\_1/USPTO.spool.p/US10791017/runat.17022005.125809.22196/app.query.faeta\_1.1358  
-DB=Published Applications NA -OPT=fastcd -SUFFIX=p2n.rnpb -MINMATCH=0.1  
-LOOPEXT=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62  
-TRANS=human40.cdt -LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100  
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0  
-MAXLEN=200000000 -USER=US10791017@cgn 1 1 879 @runat.17022005.125809.22196  
-NCPUS=6 -ICPU=3 -NO MAP -LARGQUERY -NEG SCORES=0 -WAIT -DSBLOCK=100  
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5  
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database: Published Applications\_NA:\*  
1: /cgn2\_6/prodata/1/pubpna/US07\_PUBCOMB.seq:\*  
2: /cgn2\_6/prodata/1/pubpna/PCT\_NEW\_PUB.seq:\*  
3: /cgn2\_6/prodata/1/pubpna/US06\_NEW\_PUB.seq:\*  
4: /cgn2\_6/prodata/1/pubpna/US06\_PUBCOMB.seq:\*  
5: /cgn2\_6/prodata/1/pubpna/US07\_NEW\_PUB.seq:\*  
6: /cgn2\_6/prodata/1/pubpna/PCTUS\_PUBCOMB.seq:\*  
7: /cgn2\_6/prodata/1/pubpna/US08\_NEW\_PUB.seq:\*  
8: /cgn2\_6/prodata/1/pubpna/US08\_PUBCOMB.seq:\*  
9: /cgn2\_6/prodata/1/pubpna/US09A\_PUBCOMB.seq:\*  
10: /cgn2\_6/prodata/1/pubpna/US09B\_PUBCOMB.seq:\*  
11: /cgn2\_6/prodata/1/pubpna/US09C\_PUBCOMB.seq:\*  
12: /cgn2\_6/prodata/1/pubpna/US09\_NEW\_PUB.seq:\*  
13: /cgn2\_6/prodata/1/pubpna/US10A\_PUBCOMB.seq:\*  
14: /cgn2\_6/prodata/1/pubpna/US10B\_PUBCOMB.seq:\*  
15: /cgn2\_6/prodata/1/pubpna/US10C\_PUBCOMB.seq:\*  
16: /cgn2\_6/prodata/1/pubpna/US10D\_PUBCOMB.seq:\*  
17: /cgn2\_6/prodata/1/pubpna/US10E\_PUBCOMB.seq:\*  
18: /cgn2\_6/prodata/1/pubpna/US10F\_PUBCOMB.seq:\*  
19: /cgn2\_6/prodata/1/pubpna/US10F\_NEW\_PUB.seq:\*  
20: /cgn2\_6/prodata/1/pubpna/US11\_NEW\_PUB.seq:\*  
21: /cgn2\_6/prodata/1/pubpna/US60\_NEW\_PUB.seq:\*  
22: /cgn2\_6/prodata/1/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3633	100.0	2390	9 US-09-880-107-3769	Sequence 3769, Ap
2	3633	100.0	2390	10 US-09-960-706-1081	Sequence 1081, Ap
3	3633	100.0	2390	10 US-09-873-319-717	Sequence 717, App
4	3633	100.0	2390	18 US-10-791-017A-1	Sequence 1, Appl
5	3605	99.2	2176	9 US-09-822-830A-49	Sequence 49, Appl
6	3483	95.9	2273	9 US-09-822-830A-410	Sequence 410, App
7	3277	90.2	1988	17 US-10-094-749-1074	Sequence 1074, Ap
8	1282	35.3	1824	17 US-10-439-703-58	Sequence 58, Appl
9	1281.5	35.3	1832	18 US-10-755-889-649	Sequence 649, App
10	1273.5	35.1	1939	10 US-09-919-039-322	Sequence 322, App
11	1219	33.6	763	18 US-10-425-115-182496	Sequence 182496,
12	1181	32.5	2299	14 US-10-198-846-9847	Sequence 9847, Ap
13	961	26.5	550	10 US-09-918-995-27690	Sequence 27690, A
14	868.5	23.9	568	10 US-09-918-995-9557	Sequence 9557, Ap
15	839	23.1	540	9 US-09-864-761-8001	Sequence 8001, Ap
16	774.5	21.3	505	9 US-09-864-761-7766	Sequence 7766, Ap
17	718.5	19.8	545	9 US-09-925-301-669	Sequence 669, Ap
18	684	18.8	401	10 US-09-918-995-3868	Sequence 3868, Ap
19	673	18.5	399	9 US-09-864-761-24705	Sequence 24705, A
20	631	17.4	417	9 US-09-960-352-10259	Sequence 10259, A
21	593	16.3	361	9 US-09-960-352-3851	Sequence 3851, Ap
22	496	13.7	521	9 US-09-933-797-481	Sequence 481, App
23	493	13.6	503	9 US-09-864-761-8728	Sequence 8728, Ap
24	493	13.6	503	16 US-10-029-386-4463	Sequence 4463, Ap
25	486	13.4	262	16 US-10-029-386-18163	Sequence 18163, A
26	486	13.4	1682	17 US-10-172-118-1872	Sequence 1872, Ap
27	484	13.3	1684	17 US-10-342-887-1872	Sequence 1872, Ap
28	484	13.3	1684	10 US-09-919-039-323	Sequence 323, App
29	475	13.1	253	9 US-09-864-761-25438	Sequence 25438, A
30	470.5	13.0	3061	15 US-10-101-510-602	Sequence 602, App
31	460	12.7	457	18 US-10-469-285-488	Sequence 488, App
32	423	11.6	1401	17 US-10-437-963-89443	Sequence 89443, A
33	420	11.6	1397	17 US-10-425-114-31980	Sequence 31980, A
34	420	11.6	1387	18 US-10-425-115-42134	Sequence 42134, A
35	414	11.4	483	10 US-09-918-995-2865	Sequence 2865, Ap
36	400.5	11.0	5460	10 US-10-918-715-225	Sequence 225, App
37	400.5	11.0	5460	15 US-10-171-293-67	Sequence 67, Appl
38	400.5	11.0	5460	15 US-10-301-823-32	Sequence 32, Appl
39	400.5	11.0	5460	17 US-10-257-021-71	Sequence 71, Appl
40	400.5	11.0	5460	17 US-10-172-118-428	Sequence 428, App
41	400.5	11.0	5460	17 US-10-342-887-428	Sequence 428, App
42	400.5	11.0	5460	18 US-10-474-794-225	Sequence 225, App
43	400.5	11.0	5460	18 US-10-807-308-15	Sequence 15, Appl
44	400.5	11.0	5489	16 US-10-096-534-13	Sequence 13, Appl
45	400.5	11.0	5489	17 US-10-133-937-66	Sequence 66, Appl

## ALIGNMENTS

RESULT 1  
US-09-880-107-3769  
Sequence 3769, Application US/09880107  
Patent No. US2002042981A1  
GENERAL INFORMATION:  
APPLICANT: Horne, Darci T.  
APPLICANT: Vockley, Joseph G.  
APPLICANT: Scherf, Uwe  
APPLICANT: Gene Logic, Inc.  
TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer  
FILE REFERENCE: 44921-5028-WO  
CURRENT APPLICATION NUMBER: US/09/880,107  
CURRENT FILING DATE: 2001-06-14  
PRIOR APPLICATION NUMBER: US 60/211,379  
PRIOR FILING DATE: 2000-06-14  
PRIOR APPLICATION NUMBER: US 60/237,054  
PRIOR FILING DATE: 2000-10-02  
NUMBER OF SEQ ID NOS: 3950  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 3769

```

; LENGTH 2390
;
; TYPE DNA
;
; ORGANISM: Homo sapiens
;
; FEATURES:
;
; OTHER INFORMATION: Genbank Accession No. US20020142581A1 X66959
;
US-09-860-107-3769

```

Alignment Scores:	
Pred. No.:	3,66e-253
Score:	653.00
Percent Similarity:	100.00%
Best Local Similarity:	100.00%
Query Match:	100.00%
DB:	9
Length:	2399
Matches:	656
Conservative:	0
Mismatches:	0
Indels:	0
Gaps:	0

US-10-791-017A-2 (1-656) x US-09-880-107-3769 (1-2390)

QY	1	MeAlSerThrAspIysSerThrYrSerGlnAlaAlaGlnGlnGlyYrSerAla	20
Db	44	ATGGCGTCCAGGATTAACGACTTAACCTTAACGAGCTGACGCGCAAGGCTCAAGTCT	103
QY	21	TyrThrAlaGlnProThrglnGlyYrAlaGlnThrThrglnAlaYrGlyGlnGlnSer	40
Db	104	TACACCGCCACGCGCACTCAAGGATATGACACAGACCAACCGACGATATGGCCAAACAAAGC	163
QY	41	TyrGlyThrYrGlyGlnProThrAspValSerYrThrGlnAlaGlnThrThrAlaThr	60
Db	164	TATGGAACCTATGAGACGCGCACTGATGTGACGTATACCAAGGCTCAACACATCGCAACC	223
QY	61	TyrGlyGlnThrAlaYrAlaThrSerYrGlyGlnProProThrglyYrThrPro	80
Db	224	TATGGGCAAGCGCGCTATGCACTTGTATGACAGCGCTCCACGCTGTTATACACTCCA	283
QY	81	ThrAlaProGlnAlaYrSerGlnProValGlnGlyYrGlyYThrGlyAlaYrAspThr	100
Db	284	ACGCGCCCGCAGGCAATACAGCCAGCGCTGTCCAGGGGTATGGCAGTGGCTTAATGATACC	343
QY	101	ThrThrAlaThrValThrThrThrglnAlaSerYrAlaAlaGlnSerAlaYrGlyYThr	120
Db	344	ACCACTGCTACAGTCCACACCCAGCGCTCTTATGAGCTCAATCTGACATATAGGCACT	403
QY	121	GlnProAlaYrProAlaYrGlyGlnGlnProAlaAlaThrAlaProThrArgProGln	140
Db	404	CACCTGTCTTATCAGCTATGAGGACGACGACGACGACCATGCACTTAACAAGCCGAG	463
QY	141	AspGlyAsnIysProThrGlnThrSerGlnProGlnSerSerThrGlyGlyYrAsnGln	160
Db	464	GATGGAAACAAAGCCCACTGAGACTAATCAACTCAATCAAGCAGGGGGTTTAAACAACAG	523
QY	161	ProSerIeuGlyYrGlyGlnSerAenYrSerYrProGlnValProGlySerYrPro	180
Db	524	CCAGGCTTAGAATATGACAGATACATCAATATCCCAAGTACTCGGAGACTACCCC	583
QY	181	MetGlnProValThrAlaProProSerYrProProThrSerYrSerSerThrGlnPro	200
Db	584	ATGACACCAAGTCACTGACCTCATCTCACTCTCTCAACAGGATATTCCTTAACAAGCGG	643
QY	201	ThrSerYrAspGlnSerSerYrSerGlnGlnAenYrYrGlyGlnProSerSerYr	220
Db	644	ACTAGTTATGATAGACGAGTATCTTCCAGCAACACTTAATGGGCAACCAAGACGCTAT	703
QY	221	GlyGlnGlnSerSerYrGlyGlnGlnSerSerYrGlyGlnGlnProProThrSerYr	240
Db	704	GGCACACAGATAGCTATGCTACAAACACACTATGGGCAAGACCTCCCACTAAGTTAC	763
QY	241	ProProGlnThrGlySerYrSerGlnAlaProSerGlnYrSerGlnGlnSerSer	260
Db	764	CCACCCCAACTGGAATCTTACAGCCACAGCTCCAAATCAATATAGCCAAACAAGACAGC	823
QY	261	TyrGlyGlnGlnSerSerPheArgGlnAspHisAspSerSerMetGlyValYrGlyGln	280
Db	824	TACGGGACAGAGTTCATTTCCGACAGGACCAACCCAGTACAGATGGGGTATTATGGGACG	883

QY	281	llusertgilygliphesertgilyproglilylunashrgsermetsergilyproasparnar	300
Db	884	gagrtcgagagattttccggacacgagagaaacggagacagagrtggcccttgatpaaacgg	943
QY	301	gllyaragilyaragilyglilypheasparagilygllymetseratrgilygllyaragilyglily	320
Db	944	ggcagggagagagggggatttgatctgtagagcatgacagaggtggcggggagagaga	1003
QY	321	ArgilygllymetgilyseralagilygllyuaragilyglilypheanlybproglilyProwet	340
Db	1004	CGCGGTGGAAATGGCGCACGCTGGAGAGCGAGGTGCTTCATTAAGCCTGTGGACCCATG	1063
QY	341	AspGilyglilyProAspLeuAspLeuGilyProProValAspProAspGlyAspSerAspAsn	360
Db	1064	GATAAAGACACGAATCTTGATCTTAGGCCCTCTGTAAATCCAGATGAAGACTTCGACAC	1123
QY	361	SerAlaIleTyrValGlnGlyLeuAsnAspSerValThrLeuAspAspLeuAlaAspPhe	380
Db	1124	AGTGCATTTTGTACAGAGATTAAATGACAGTGTGACTCTAGATGATCTGGCAGACTTC	1183
QY	381	PheIysGlnCybGilyValIallyMetAsnlysaTrgThrgIyGlnProwetIleHsIle	400
Db	1184	TTTAAAGCAGTGTGGGTGTGTTAAGATGAACAAGAACTGGGCAACCATGATCAACATC	1243
QY	401	TyrlleAspIlybGlyThTrgIlybPProIyAspAlaThrValSerTyrGlyAsnPro	420
Db	1244	TACCTGACAGAGAAACAGAAACCCAAAGGCATGCCACAGTCTCTTAAGAAACCA	1303
QY	421	ProThrAlaIlybAlaIlyValGlyTrpPheAspGilybAspPheGlnGlySerIlyLeu	440
Db	1304	CCCATCTGCAGAGCTGCCGTGGATGTGTGATGGAAAGATTTTCAAAGGAGCAAACTT	1363
QY	441	LyvalSerleuAlaArgIlybPProProwetAsnSerMetAArgIlyGlyLeuProPro	460
Db	1364	AAAGTCTCCCTTGTCTCGAAGAAACCTCCAATGAACAGTATGCGGGGTGGCTGCCACC	1423
QY	461	ArgGilyglilyaragilybmetProProProProwetAsnArgIlyglilyProglilyglilyPro	480
Db	1424	CGTAAAGGACAGAGGACATGCCACCACTCGTGGAGGTCCAGAGGCCCCAGAGGTCTT	1483
QY	481	GlyglilyProwetGilyarGmetGlyGlyAArgGilyglilybAspAArgGilyglilyPheProPrtArg	500
Db	1484	GGGGGACCCATGGGTCCGATGGAGGCCGTGGAGAGATGAGAGGCTTCCCTCCAAAG	1543
QY	501	GlybProAArgIlySerAArgIlyAsnProSerGilyglilyGlyAsnValGlnHsAArgAlaGly	520
Db	1544	GGAACCCGGGGTTCCTCGAGGGAAACCTCTGGAGAGAGAAACGTCCAGACCGAGCTGGA	1603
QY	521	AspTrpGlnCybProAsnProIyCybGlybAsnGlnAsnPheAlaTrpAlyThGlybCy	540
Db	1604	GACTGGCAGGTGCTCCCAATCCGGGTGTGGAAACAAGAACTTCGCTGGAGAAACAGAGTGC	1663
QY	541	AsnGlnCybIlybAlaPProIyPProGlnGlyPheLeuProProProwetProProGly	560
Db	1664	AACCAAGTGAAGGCCCAAGACCTGAAGGCTTCTCCCGCACCTTTCGCCCCGGGT	1723
QY	561	GlybAspAArgIlyAArgIlyglilyProglilyGlyMetAArgIlyglilyAArgIlyglilyLeuMetAsp	580
Db	1724	GGTATGTGGCAGAGGTGGCCCTGTGGCATGCGGGGAGGAAAGAGGTGGCTCATGGAT	1783
QY	581	ArgGilyglilyProglilyglilyMetPheAArgIlyglilyAArgIlyglilybAspAArgIlyglilyPheArg	600
Db	1784	CGTGTGTGTCCCGGTGGAAATGTTCAAGAGTGGCGCTGTGTGAGACAAAGGTGTGCTTCGT	1843
QY	601	GlyglilyaragilybmetAspAArgIlyglilyPheGlyglilyglilyaragagilyglilyProglilyglily	620
Db	1844	GGTGGCCGGGACATGACCAAGAGGTGCTTTGTGGAGAAACAAGAGTGGCCCTGGGGGG	1903
QY	621	ProProGlyProLeuMetGlnGlnMetGlyglilyAArgAArgIlyglilyAArgGilyglilyProglily	640
Db	1904	CCCCCTGGACCTTTGATGGAAACAGATGGAGAGAAAGAGAGGACCTGTGAGACCTGGA	1963
QY	641	LybMetAspIlybGlylunHsAArgGlnIlyAArgIlybAArgAArgProTyr	656



Db 1964 AAAATGATTAAGCGACGACCGCTCAGAGCGCAGAGATGGCCCTTAC 2011  
RESULT 2  
US-09-960-706-1081  
Sequence 1081, Application US/09960706  
Publication No. US20030134280A1  
GENERAL INFORMATION:  
APPLICANT: Munger, William E.  
TITLE OF INVENTION: Identifying Drugs for and diagnosis of Benign Prostatic Hyperplasia  
FILE REFERENCE: 44921-5029-01US  
CURRENT APPLICATION NUMBER: US/09/960,706  
PRIORITY FILING DATE: 2001-09-24  
PRIORITY FILING DATE: 2000-08-07  
PRIORITY FILING DATE: 2000-08-07  
PRIORITY FILING DATE: 2001-06-05  
NUMBER OF SEQ ID NOS: 1124  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 1081  
LENGTH: 2390  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
OTHER INFORMATION: Genbank Accession No. US20030134280A1 X66899  
US-09-960-706-1081  
Alignment Scores:  
Pred. No.: 3,666-253 Length: 2390  
Score: 3633.00 Matches: 656  
Percent Similarity: 100.008 Conservative: 0  
Best Local Similarity: 100.008 Mismatches: 0  
Query Match: 100.004 Indels: 0  
Gaps: 0  
US-10-791-017a-2 (1-656) x US-09-960-706-1081 (1-2390)  
QY 1 MetAlaSerThrAspTyrSerThrTyrSerGlnAlaAlaGlnGlnGlnTyrSerAla 20  
Db 44 ATGGCGGCCAGCGATTACATCTAATACCAAGCTCAGCGAGCGGCTACAGTGTCT 103  
QY 21 TTTThAlaGlnProthGlnGlnGlnTyrAlaGlnThrGlnAlaTyrGlnGlnGlnSer 40  
Db 104 TACACCGCCAGCGCCACTCAAGATATGCAAGCAAGCCAGCGATATGAGCAAGC 163  
QY 41 TTTGlyThrTyrGlnGlnProthAspValSerTyrThrGlnAlaGlnThrAlaThr 60  
Db 164 TATGGAACCTATGAGACGCGCACTGATGTCACTATCCAGGCTCAGACCTGCAACC 223  
QY 61 TTTGlyGlnThrAlaTyrAlaThrSerTyrGlyGlnProthGlnTyrThrThrPro 80  
Db 224 TATGGGACAGCCCGCTATGCACTTCTTATGACAGCTCCCACTGGTTATTAATCTCA 283  
QY 81 ThrAlaProGlnAlaTyrSerGlnProValGlnGlnTyrGlnTyrAlaTyrAspThr 100  
Db 284 ACTGCCCCCAGGCGATACAGCCAGCTGTCCAGGAGATGAGCACTGTGTCTTATGATACC 343  
QY 101 ThrThrAlaThrValThrThrGlnAlaSerTyrAlaAlaGlnSerAlaTyrGlyThr 120  
Db 344 ACAACGTCTACATCACTCAACCAAGCGCTCTCTTATGACCTAGTGTGATATGGCACT 403  
QY 121 GlnProAlaTyrProAlaTyrGlnGlnGlnProAlaAlaThrAlaProthThrProGln 140  
Db 404 CACCCGCTTATCCAGCTATGGGACAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 463  
QY 141 AspGlyAsnTyrProthGlnTyrSerGlnProGlnSerSerThrGlnGlnTyrAspGln 160  
Db 464 GATGGAAGAACAGCCCACTGAGATGATCACTCAATCTATGACAGAGGAGGATTTCAACAG 523  
QY 161 ProSerLeuGlnTyrGlnGlnSerAsnTyrSerTyrProGlnValProGlySerTyrPro 180  
Db 524 CCAAGCTAGATATGAGCAGAGATTAAGTTATCCAGATGATCTTGGAGACTTACCCC 583

QY 181 MetGlnProValThrAlaProProSerTyrProProthSerTyrSerSerThrGlnPro 200  
Db 584 ATGACAGCACTCACTCACTCACTCACTCACTCACTCACTCACTCACTCACTCACTCA 643  
QY 201 ThrSerTyrAspGlnSerSerTyrSerGlnAlaThrTyrGlnGlnProSerSerTyr 220  
Db 644 ACTAGTTATATATAGAGCACTTCTCTCAGAGAACTATATGAGCAAGCAAGCAAGCA 703  
QY 221 GlnGlnGlnSerSerTyrGlnGlnGlnSerSerTyrGlnGlnGlnProProthSerTyr 240  
Db 704 GACAGAGAGTATGATATGATATGATATGATATGATATGATATGATATGATATGATAT 763  
QY 241 ProProGlnThrGlnSerTyrSerGlnAlaProSerGlnTyrSerGlnGlnSerSer 260  
Db 764 CCAAGCCCAAGTCACTCACTCACTCACTCACTCACTCACTCACTCACTCACTCACT 823  
QY 261 TTTGlyGlnGlnSerSerPheArgGlnAspPheProSerSerMetGlyValTyrGlyGln 280  
Db 824 TACGGGACAGAGCTTCACTTCCAGAGACCACTCCAGTGCATGGGTGTTATGGGACG 883  
QY 281 GluSerGlyGlyPheSerGlyProGlyGluAsnArgSerMetSerGlyProAspAsnArg 300  
Db 884 GAGTCTGAGAGATTTTCCGAGCCAGAGAGAACCGAGATGATGAGCTCCGATTAACCG 943  
QY 301 GlnArgGlyArgGlyGlyPheAspArgGlyGlyMetSerArgGlyGlyValArgGly 320  
Db 944 GGCAGGAGAAAGAGGGGATTTGATCGTGAAGGATGAGAGAGTGGGCGGAGAGAGGA 1003  
QY 321 ArgGlyGlyMetGlySerAlaGlnGlyGluArgGlyGlyPheAsnTyrProGlyProMet 340  
Db 1004 CGCGGTGAGATGGGACGCTGAGAGACAGAGTGGCTTCAATTAAGCTTGTGAGACCATG 1063  
QY 341 AspGlnGlyProAspLeuAspLeuGlyProProValAspProAspGluAspSerAsp 360  
Db 1064 GATGAAGAGCAAGATCTTGAATAGGCGCTCTGTAGATCAAGATGAAGCTGCAAC 1123  
QY 361 SerAlaLeuTyrValGlnGlyLeuAsnAspSerValThrLeuAspLeuAlaAspPhe 380  
Db 1124 AGTCAATTTATTAACAAGATTAATGAACAGTGAATCTTAATATGATGAGCACTTC 1183  
QY 381 PheLeuGlnGlyGlyValValLysMetAsnLysArgThrGlyGlnProMetLysLys 400  
Db 1184 TTTAAGCATGTGGGTGTTTAAAGTGAACAAGAACTGGGCAACCATGATTCACATC 1243  
QY 401 TTTLeuAspLysGlnThrGlyLysProLysGlyAspAlaThrValSerTyrGluAspPro 420  
Db 1244 TACCTGACAGAGAAACAGGAAGCCCAAGGAGTCCACAGTGTCTTATGAAGCCCA 1303  
QY 421 ProThrAlaLysAlaAlaValGluTTPPheAspGlyLysAspPheGlnGlySerLysLeu 440  
Db 1304 CCCAGTCCCAAGGCTCCGCGAATGTTGAAGGAAAGATTTTCAAGGAGCAAACTT 1363  
QY 441 LysValSerLeuAlaArgLysLysProProMetAsnSerMetArgGlyGlyLeuProPro 460  
Db 1364 AAAGTCTCCCTGCTGGAAGAGCTTCCAGTAAGATGATGAGGAGTGTCTGCAACC 1423  
QY 461 ArgGlnGlyArgGlyMetProProProLeuArgGlyGlyProGlyGlyProGlyPro 480  
Db 1424 CGTGAAGGAGAGGATGCAACCACTCCGAGAGTCCAGAGAGCCAGAGAGTCT 1483  
QY 481 GlyLysProMetGlyArgMetGlyLysArgGlyGlyAspArgGlyGlyPheProProArg 500  
Db 1484 GGGGAGACCATGGGTGCAATGGAGGCGCGTGAAGATGAGAGAGGCTTCCCTCCAGA 1543  
QY 501 GlyProArgGlySerArgGlyAsnProSerGlyGlyLysAsnValGlnLysArgAlaGly 520  
Db 1544 GAGCCCCGGGGTTCGAGAGGAACCCCTGAGAGAGGAACCTCCAGCAAGAGCTGGA 1603  
QY 521 AspTyrGlnCysProAsnProGlyCysGlyAsnGlnAsnPheAlaThrArgThrGluCys 540  
Db 1604 GACTGCAAGTGTCCCAATCCGGGTTGTGGAACCAACAATCTTCCCTGAGGAACAGAGTGC 1663

QY 541 AsnGlnCysIysAlaProLysProGluGlyPheLeuProProPheProProGly 560  
 Db 1664 AACCAAGTAAAGGCCCAAAAGCCTGAAAGCTTCTCCCGCACTTTCGGCCCCGGGT 1723  
 QY 561 G1YAAspArgGlyArgGlyGlyProGlyGlyMetArgGlyGlyArgGlyGlyLeuMetAsp 560  
 Db 1724 GGTATGATGAGCAGAGCTGGCCCTGTGTCATCGCGGAGAGAGAGCTGCTCATGAT 1783  
 QY 581 ArgGlyGlyProGlyGlyMetPheArgGlyGlyArgGlyGlyAspArgGlyGlyPheArg 600  
 Db 1784 CGTGTGTATCCCGGTGAAAGTTTCAAGAGTGGCCGTGTGAGACAAAGGTGGCTTCGT 1843  
 QY 601 G1Yg1YArgGlyMetAspArgGlyGlyPheGlyGlyGlyArgArgGlyGlyProGlyGly 620  
 Db 1844 GGTGGCCGGGGGCAAGCAGAGTGGCTTGTGTGAGAGAGACAGAGTGGCCCTGGGGGG 1903  
 QY 621 ProProGlyProLeuMetGluGluMetGlyGlyArgArgGlyGlyArgGlyGlyProGly 640  
 Db 1904 CCCCTGACCTTGTGAGAAAGATGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1963  
 QY 641 LysMetAspLysGlyGlyIleArgGluArgArgArgArgArgArgArgArgArgArg 656  
 Db 1964 AAAATGATAAAGCGAGCAGCCGTCAGAGCGAGAGATCGGCCCTTAC 2011

## RESULT 3

US-09-873-319-717  
 : Sequence 717, Application US/09873319A  
 : Publication No. US20030134324A1  
 : GENERAL INFORMATION:  
 : APPLICANT: Mungier, William E.  
 : APPLICANT: Kulkarni, Prakash  
 : APPLICANT: Geitzenberg, Robert H.  
 : APPLICANT: Waga, Iwao  
 : APPLICANT: Yamamoto, Jun  
 : TITLE OF INVENTION: Identifying Drugs for and Diagnosis of Benign Prostatic  
 : FILE REFERENCE: 44921-5029-US  
 : CURRENT APPLICATION NUMBER: US/09/873,319A  
 : EARLIER FILING DATE: 2001-06-05  
 : EARLIER APPLICATION NUMBER: US 60/223,323  
 : NUMBER OF SEQ ID NOS: 755  
 : SOFTWARE: PatentIn Ver. 2.1  
 : SEQ ID NO 717  
 : LENGTH: 2390  
 : TYPE: DNA  
 : ORGANISM: Homo sapiens  
 : FEATURE:  
 : OTHER INFORMATION: Genbank Accession No. US20030134324A1 X66899  
 : US-09-873-319-717

## Alignment Scores:

Pred. No.: 3,666-253 Length: 2390  
 Score: 3633.00 Matches: 656  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: 10 Gaps: 0

US-10-791-017a-2 (1-656) x US-09-873-319-717 (1-2390)

QY 1 MetAlaSerThrAspTyrSerThrTyrSerGlnAlaAlaAlaGlnGlnGlyTyrSerAla 20  
 Db 44 ATGGCGTCCACGAGATTACAGTACTATACCAAGCTCAGCGCAGCGGCTACAGTGTCT 103  
 QY 21 TyrThrAlaGlnProThrglnGlyTyrAlaGlnThrThrglnAlaTyrGlyGlnGlnSer 40  
 Db 104 TACACCGCCAGCCCACTCAAGATATGACAGACCAACCAAGGATATGGCAACAAAGC 163  
 QY 41 TyrGlyThrTyrGlyGlnProThraAspValSerTyrThrGlnAlaGlnThrThraAlaThr 60  
 Db 164 TATGGAACCTATGACAGCCCACTGATGTCACTATACCAAGGCTCAGACCACTGCAACC 223

QY 61 TyrGlyGlnThrAlaTyrAlaThrSerTyrGlyGlnProProThrglyTyrThrThrPro 80  
 Db 224 TATGGCAGACCGCCCTATGCAACTTATGACAGCCCTCCACTGTATATATCTCA 283  
 QY 81 ThrAlaProGlnAlaTyrSerGlnProValGlnGlyTyrGlyThrglyAlaTyrAspThr 100  
 Db 284 ACTGCCCCAGCAGACACAGCCCTCTATGACGCTCAAGTGTGATATGATACC 343  
 QY 101 ThrThrAlaThrValThrThrThrglnAlaSerTyrAlaAlaGlnSerAlaTyrGlyThr 120  
 Db 344 ACCACGTCTACAGTCAACACACAGCCCTCTATGACGCTCAAGTGTGATATGACT 403  
 QY 121 GlnProAlaTyrProAlaTyrGlyGlnGlnProAlaAlaThrAlaProThraProGln 140  
 Db 404 CAGCCGTCTATACAGCTATGAGGAGAGCCAGCAGCAGCAGCAGCAGCAGCAGCAG 463  
 QY 141 AspGlyAsnLysProThrglyThrSerGlnProGlnSerSerThrGlyGlyTyrAsnGln 160  
 Db 464 GATGGAACAGCCCACTGACACTAGTCAACTCAATCTAGCACAGGGGGTTCAACAG 523  
 QY 161 ProSerLeuGlyTyrGlyGlnSerAsnTyrSerTyrProGlnValProGlySerTyrPro 180  
 Db 524 CCAAGCCTAGATATGACAGAGTAACAGTATCACTATCCAGATACCTGGAGACTACCC 583  
 QY 181 MetGlnProValThrAlaProProSerTyrProProThrglySerSerThrGlnPro 200  
 Db 584 ATGCACACGACTACCTGACCTCACTCACTCACTCACTCACTCACTCACTCACTCA 643  
 QY 201 ThrSerTyrAspGlnSerSerTyrSerGlnGlnThrThrglyGlnProSerSerTyr 220  
 Db 644 ACTAGTATGATCAGACAGCTTACTCTCAGCAGAACCACTATGGCAGACAGCACTAT 703  
 QY 221 GlnGlnGlnSerSerTyrGlyGlnGlnSerSerTyrGlyGlnGlnProProThrglyTyr 240  
 Db 704 GACACACAGAGTACTATGATGATCAACAGAGTATGAGGAGAGAGCTCCACTAGTAC 763  
 QY 241 ProProGlnThrglySerTyrSerGlnAlaProSerGlnTyrSerGlnGlnSerSer 260  
 Db 764 CACGCCCAACAGTGGATCTTACAGCCAGCAAGCTCCAGTCAATATAGCAGACAGCAG 823  
 QY 261 TyrGlyGlnGlnSerSerPheArgGlnAspHisProSerSerMetGlyValTyrGlyGln 280  
 Db 824 TACGGGCAAGAGTTCATTCGACAGAGCAACCCAGATGACATGGGTGTTATGGGCA 883  
 QY 281 GluSerGlyGlyPheSerGlyProGlyGluAsnArgSerMetSerGlyProAspAsnArg 300  
 Db 884 GAGCTGAGAGATTTCCGAGCCAGAGAGAAACGAGCATGAGTGGCCCTGATTAACCG 943  
 QY 301 G1YArgGlyArgGlyGlyPheAspArgGlyGlyMetSerArgGlyGlyArgGlyGly 320  
 Db 944 GGCAGGGGAAGAGGGGATTTGATCGTGAAGCATGACAGAGTGGGGGGGAGAGGA 1003  
 QY 321 ArgGlyGlyMetGlySerAlaGlyGluArgGlyGlyPheAsnLysProGlyGlyProMet 340  
 Db 1004 CGCGGTGATGATGGCGCCCTGAGAGCAGAGGTGGCTTCAATTAAGCTGTGAGACCAATG 1063  
 QY 341 AspGlyGlyProAspLysAspLeuGlyProProValAspProAspGlyLysAspAspAsn 360  
 Db 1064 GATGAAGACCAAGATTTATCTTAGCCCTCTGTGATCCAGATAGAGCTTTCACAC 1123  
 QY 361 SerAlaIleTyrValGlnGlyLeuAsnAspSerValThrLeuAspAspLeuAlaAspPhe 380  
 Db 1124 AGTGCAATTTATGATCAAGATTTAATGACAGTGTACTCTGATGATCTGGCAGACTTC 1183  
 QY 381 PheLeuGlnCysGlyValValLysMetAsnLysArgThrGlyGlnProMetIleHisIle 400  
 Db 1184 TTTAAGCACTGTGGGCTTGTAAATGAACAAGAGAACCTGGGCAACCAATGATCCACATC 1243  
 QY 401 TyrLeuAspLysGlnThrGlyLysProLysGlyAspAlaThrValSerTyrGlyLysAspPro 420  
 Db 1244 TACCTGGAACAGGAACAGAAAGCCCAAGGAGATGCCACAGTGTCTATGAAAGCCCA 1303  
 QY 421 ProThrAlaLysAlaAlaValGluTyrPheAspGlyLysAspPheGlnGlySerLysLeu 440



QY 341 AspGluGlyProAspLeuAspLeuGlyProProValAspProAspGlyAspSerAspAsn 360  
 Db 1064 GATGAAAGACCAAGATCTTGATCTAGGCCCACTTAACTCAAGTAGAGCTCTGACAC 1123  
 QY 361 SerAlaIleTyrValGlnGlyLeuAsnAspSerValThrLeuAspLeuAlaAspPhe 380  
 Db 1124 AGTGCAATTTATGTAACAAGATTAAATGACAGTGTGACTGATGATGATCTGGACGACTTC 1183  
 QY 381 PheIleGlnCysGlyValValIleValMetAsnLysArgThrGlyGlnProMetIleHisIle 400  
 Db 1184 TTTAAGAGAGTGGGTGTGTATGATGAAACAAGAACTGGGCAACCAATGATCCACATC 1243  
 QY 401 TyrLeuAspLysGluThrGlyLysProLysGlyAspAlaThrValSerTyrGluAspPro 420  
 Db 1244 TACCTGACAAAGAAACAAGAAAGCCAAAGGCAATCCACAGTCTCTTGAAGACCA 1303  
 QY 421 ProThrAlaLysAlaAlaValGluTrpPheAspGlyLysAspPheGlnGlySerLysLeu 440  
 Db 1304 CCACCTCCAAAGGCTCCGTGGAATGGTTGATGGGAAAGATTTTCAAGGGAACAACTT 1363  
 QY 441 LysValSerLeuAlaGlyLysProProMetAsnSerMetArgGlyGlyLeuProPro 460  
 Db 1364 AAAGTCTCCCTTGCTCGAAAGAACCTCCATGAAACAGTATGCGGGTGTCTGCCACCC 1423  
 QY 461 ArgGluGlyLysArgLysMetProProProLeuAspGlyGlyLysProGlyGlyPro 480  
 Db 1424 CGTAGAGGACAGAGCATGCTCCACCACTCCGTGAGAGTCCAGAGGCCCAAGAGTCTCT 1483  
 QY 481 GlyLysProMetGlyLysArgMetGlyLysArgGlyLysAspArgGlyLysPheProProArg 500  
 Db 1484 GGGGGAACCAATGGGTCCGATGGAGGCGGTGAGAGATAGAGGAGGCTTCCCTCCAAAG 1543  
 QY 501 GlyProArgGlySerArgGlyAsnProSerGlyGlyLysValAsnValGlnHisArgAlaGly 520  
 Db 1544 GGAACCCGGGGTCTCCAGAGGAACCTCTGAGAGAGAAACGTCACGACGAGCTGGA 1603  
 QY 521 AspTrpGlnCysProAsnProGlyLysGlyLysAsnGlnAsnPheAlaTPAqThrGluCys 540  
 Db 1604 GACTGGAGAGTCTCCCAATCCGGGTGTGGAACCAAGACTTCCGCTGGAAGAACAGACTGC 1663  
 QY 541 AsnGlnCysAlaLysAlaProLysProGluGlyLysPheLeuProProPheProProGly 560  
 Db 1664 AACCAAGTAAAGGCCCAAGAGCTGAAAGCTTCTCCGCGCACCTTCCGCGCGGGGT 1723  
 QY 561 GlyAspArgGlyLysArgGlyLysProGlyGlyLysMetArgGlyLysArgGlyLysLeuMetAsp 580  
 Db 1724 GGTGATCGTGCAGAGGTGGCTGTGTGTCATGCGGAGAGAAAGGTGGCTTCATGAT 1783  
 QY 581 ArgGlyLysProGlyLysMetPheArgGlyLysArgGlyLysAspArgGlyLysPheArg 600  
 Db 1784 CGTGTGTCTCCGCTGGAATGTTCAAGAGTGGCCGTGTGGAAGCAAGAGTGGCTTCCGT 1843  
 QY 601 GlyLysArgGlyLysMetAspArgGlyLysPheGlyLysLysArgGlyLysProGlyLys 620  
 Db 1844 GGTGCGCGGGCATGGAACGAGGTGCTGTGTGAGAAAGAAAGAGAGAGAGAGAGAGAGAGAG 1903  
 QY 621 ProProGlyProLeuMetGluGlnMetGlyLysArgArgGlyLysArgGlyLysProGly 640  
 Db 1904 CCCCTGACCTTTGAGTGAAGACAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1963  
 QY 641 LysMetAspLysGlyLysLysArgGlnGlyLysArgArgAspArgProTyr 656  
 Db 1964 AAAATGATAAAGCGAGACCGTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2011

; APPLICANT: Fechtel, Kim  
 ; APPLICANT: Agostino, Michael J.  
 ; APPLICANT: Howes, Steven H.  
 ; APPLICANT: Resnick, Richard J.  
 ; APPLICANT: Gulikota, Kamalakari  
 ; APPLICANT: Graham, James R.  
 ; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING NOVEL SECRETED PROTEINS  
 ; FILE REFERENCE: GIN 6402  
 ; CURRENT APPLICATION NUMBER: US/09/822, 830A  
 ; PRIOR FILING DATE: 2001-03-29  
 ; PRIOR APPLICATION NUMBER: 60/195, 604  
 ; NUMBER OF SEQ ID NOS: 631  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO 49  
 ; LENGTH: 2176  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 ; US-09-822-830A-49

Alignment Scores:  
 Pred. No.: 3,53e-251 Length: 2176  
 Score: 3605.00 Matches: 655  
 Percent Similarity: 99.85% Conservative: 0  
 Best Local Similarity: 99.85% Mismatches: 1  
 Query Match: 99.23% Indels: 1  
 DB: 9 Gaps: 0

US-10-791-017A-2 (1-656) x US-09-822-830A-49 (1-2176)

QY 1 MetAlaSerThrAspTyrSerThrTyrSerGlnAlaAlaGlnGlnGlyTyrSerAla 20  
 Db 2160 ATGGGCTCCACGGATTAAGTACAGTACCTATGACCAAGTGCAGCGGAGGCTCAAGTCT 2101  
 QY 21 TyrThrAlaGlnProThrGlnGlyTyrAlaGlnThrThrGlnAlaTyrGlyGlnSer 40  
 Db 2100 TACACCGCCGAGGCCCTCAAGATATGCAAGAACCAACCGAGCATATGGCAACAAAGC 2041  
 QY 41 TyrGlyThrTyrGlyGlnProThrAspValSerTyrThrGlnAlaGlnThrThrAlaThr 60  
 Db 2040 TATGAACTATGAGAGAGCCCACTATATGCTATGATATCCAGGCTATAGACCTGCAACC 1981  
 QY 61 TyrGlyGlnThrAlaTyrAlaThrSerTyrGlyGlnProProThrGlyTyrThrThrPro 80  
 Db 1980 TATGGCAGACCGGCTATGCACTTCTTATGAGCAGCTCCCACTGCTTATATCTCA 1921  
 QY 81 ThrAlaProGlnAlaTyrSerGlnProValGlnGlyTyrGlyThrGlyAlaTyrAspThr 100  
 Db 1920 ACTGCCCGCCGAGCATACAGCCGCTGTCCAGGGGTATGGCACTGTGTCTATATGATACC 1861  
 QY 101 ThrThrAlaThrValThrThrThrGlnAlaSerTyrAlaAlaGlnSerAlaTyrGlyThr 120  
 Db 1860 ACCACTGCTACACTCACCAACCAAGGCTCTCTATGAGCTCAAGTGTGATATGGCACT 1801  
 QY 121 GlnProAlaTyrProAlaTyrGlyGlnGlnProAlaAlaThrAlaProThrArgProGln 140  
 Db 1800 CAGCGCTTATTCACAGCTATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1742  
 QY 141 AspGlyAsnLysProThrGluThrSerGlnProGlnSerSerThrGlyLysTyrAsnGln 160  
 Db 1741 GATGAAACAAAGCCCACTGAGACTGATCACTCACTTACAGCAAGAGGGGTATCAACAG 1682  
 QY 161 ProSerLeuGlyTyrGlyLysIleAsnTyrSerTyrProGlnValaProGlySerTyrPro 180  
 Db 1681 CCCAGCTAGATATGAGAGAGATACAGTATATCCCAAGTATCTGAGACTACCTCC 1622  
 QY 181 MetGlnProValThrAlaProProSerTyrProProThrSerTyrSerSerThrGlnPro 200  
 Db 1621 ATGCACGAGTACATCTCATCTCATCTCCCTCCACAGCATATCTCTTACAGAGCGG 1562  
 QY 201 ThrSerTyrAspGlnSerSerTyrSerGlnGlnMetThrTyrGlyGlnProSerSerTyr 220  
 Db 1561 ACTAGTTATGATCAGAGCAGTACTCTCAGAGAACCACTATGGGAGACCGAGCAGCTAT 1502

RESULT 5

US-09-822-830A-49/c  
 ; Sequence 49, Application US/09822830A  
 ; Patent No. US20020142952A1

; GENERAL INFORMATION:

; APPLICANT: Genetics Institute, Inc.

; APPLICANT: Wong, Gordon G.

; APPLICANT: Clark, Hilary

```

QY 221 G1YGLNGINSErserTYrG1YGLNGINSErserTYrG1YGLNGINSErserTYr 240
Db 1501 GGACAGAGAGTACTGATGTCACAAAGAGCTATGGCGAGCCCTCCACATGTTAC 1442
QY 241 ProProG1InThrG1SerTYrSerG1nAlaProSerG1nTYrSerG1nInSErSer 260
Db 1441 CCACCCCAACTGGATCTTACAGCCAAAGCTCCAAATGTCATATACCAACAGAGAGAGC 1382
QY 261 TYrG1YGLNGINSErserPheArG1nAph1aProSerSerMetG1YalTYrG1YGLn 280
Db 1381 TACGGGAGAGAGAGTCTTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1322
QY 281 GluSerG1YGLNPhSerG1YProG1YGLNPhArG1nArG1nArG1nArG1nArG1n 300
Db 1321 GAGCTGAGAGATTTTCCGAGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1262
QY 301 G1YArG1YArG1YGLNPhArPArG1YGLNPhArPArG1YGLNPhArPArG1YGLY 320
Db 1261 GGCGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1202
QY 321 ArG1YGLNPhArG1YArG1YGLNPhArG1YGLNPhArG1YGLNPhArG1YGLY 340
Db 1201 CGCGGTGAGATGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1142
QY 341 ArPArG1YGLNPhArPArG1YGLNPhArPArG1YGLNPhArPArG1YGLNPhAr 360
Db 1141 GATGAGAGAGAGAGATCTTGAATCTTGAATCTTGAATCTTGAATCTTGAATCT 1082
QY 361 SerAla1eTYrV1a1G1NGLNPhArPArG1YGLNPhArPArG1YGLNPhArPAr 380
Db 1081 AGTGCATTTATGATGACAGAGATTTAAAGAGATGATGATGATGATGATGATGAT 1022
QY 381 PheArG1NPhArG1YGLNPhArG1YGLNPhArG1YGLNPhArG1YGLNPhAr 400
Db 1021 TTTAAGAGAGTGGGAGTGTGTTAAGATGACAGAGAGAGAGAGAGAGAGAGAG 962
QY 401 TYrLeuArPArG1YGLNPhArPArG1YGLNPhArPArG1YGLNPhArPArG1Y 420
Db 961 TACCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 902
QY 421 ProThrAlaYsaAlaVala1G1uTPPhArPArG1YGLNPhArPArG1YGLNPh 440
Db 901 CCACCTCCAAAGGCTTCCGAGAGAGATTTGATGAGAGAGATTTTCAAGAGAGAG 842
QY 441 LyseValSerLeuAlaArG1YGLNPhArPArG1YGLNPhArPArG1YGLNPhAr 460
Db 841 AAGATCTCCCTTGTCTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 782
QY 461 ArG1YGLNPhArG1YGLNPhArPArG1YGLNPhArPArG1YGLNPhArPArG1 480
Db 781 CGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 722
QY 481 G1YGLNPhArG1YGLNPhArPArG1YGLNPhArPArG1YGLNPhArPArG1Y 500
Db 721 GGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 662
QY 501 G1YProArG1YGLNPhArPArG1YGLNPhArPArG1YGLNPhArPArG1YGL 520
Db 661 GGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 602
QY 521 AsPArG1YGLNPhArPArG1YGLNPhArPArG1YGLNPhArPArG1YGLNPh 540
Db 601 GACTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 542
QY 541 AAGG1NPhArPArG1YGLNPhArPArG1YGLNPhArPArG1YGLNPhArPAr 560
Db 541 AAGG1NPhArPArG1YGLNPhArPArG1YGLNPhArPArG1YGLNPhArPAr 482
QY 561 G1YArPArG1YGLNPhArPArG1YGLNPhArPArG1YGLNPhArPArG1YGL 580
Db 481 GGTATATGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 422

```

```

QY 581 ArG1YGLNPhArPArG1YGLNPhArPArG1YGLNPhArPArG1YGLNPhAr 600
Db 421 CGTGGTGGTCCGATGAGAGATTTCAAGAGTGGCGGTGGAGAGAGAGAGAGAG 362
QY 601 G1YGLNPhArG1YGLNPhArPArG1YGLNPhArPArG1YGLNPhArPArG1Y 620
Db 361 GTGGCGGGGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 302
QY 621 ProProG1YProLeuArG1YGLNPhArG1YGLNPhArG1YGLNPhArG1YGL 640
Db 301 CCCCTGAGAGCTTGTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 242
QY 641 LYMeArPArG1YGLNPhArPArG1YGLNPhArPArG1YGLNPhArPArG1Y 656
Db 241 AAAATGATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 194

RESULT 6
US-09-822-830A-410/c
: Sequence 410. Application us/09822830A
: Patent No. US20020142952A1
: GENERAL INFORMATION:
: APPLICANT: Genetics Institute, Inc.
: APPLICANT: Wong, Gordon G.
: APPLICANT: Clark, Hilary
: APPLICANT: Fectel, Kim
: APPLICANT: Agostino, Michael J.
: APPLICANT: Howes, Steven H.
: APPLICANT: Resnick, Richard J.
: APPLICANT: Gulukota, Kamalakart
: APPLICANT: Graham, James R.
: TITLE OR INVENTION: POLYNUCLEOTIDES ENCODING NOVEL SECRETED PROTEINS
: FILE REFERENCE: GIN 6402
: CURRENT APPLICATION NUMBER: US/09/822,830A
: CURRENT FILING DATE: 2001-03-29
: PRIOR APPLICATION NUMBER: 60/195,604
: PRIOR FILING DATE: 2000-04-06
: NUMBER OF SEQ ID NOS: 631
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 410
: LENGTH: 2273
: TYPE: DNA
: ORGANISM: Homo sapiens
US-09-822-830A-410

Alignment Scores:
Pred. No.: 2,36e-242 Length: 2273
Score: 3483.00 Matches: 636
Percent Similarity: 97.10% Conservative: 0
Best Local Similarity: 97.10% Mismatches: 1
Query Match: 95.87% Indels: 18
DB: Gaps: 2

US-10-791-017A-2 (1-656) x US-09-822-830A-410 (1-2273)
QY 2 AlAserThrArPArG1YGLNPhArPArG1YGLNPhArPArG1YGLNPhArPAr 21
Db 2263 GCGTCCACGAGATTAAGATGATGATGATGATGATGATGATGATGATGATGAT 2204
QY 22 ThrAlaG1nProThrG1NGLNPhArPArG1YGLNPhArPArG1YGLNPhArPAr 41
Db 2203 ACCGCCAGCCCACTCAAGATATGACAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2144
QY 42 G1YThrTYrG1NGLNPhArPArG1YGLNPhArPArG1YGLNPhArPArG1YGL 61
Db 2143 GGAAGCTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2084
QY 62 G1YGLNPhArPArG1YGLNPhArPArG1YGLNPhArPArG1YGLNPhArPAr 81
Db 2083 GGGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2024
QY 82 AlAProG1nAlaTYrSerG1nProVala1G1NGLYrG1YThrG1YalTYrArPAr 101
Db 2023 GCCCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1964

```

QY 102 ThrAlaThrValThrThrThrGlnAlaSerTyrAlaIaGlnSerAlaTyrGlyThrGln 121  
 Db 1963 AGGCTACAGTACACACACACAGGCTCTATCAGCTCAGTGTGATATGAGCACTCAG 1904  
 QY 122 ProAlaTyrProAlaTyrGlyGlnGlnProAlaAlaThrAlaProThrAaGProGlnAsp 141  
 Db 1903 CTGCTTATCCAGCTATGAGGACACAGCAGACACCTGACCTTACAGACCGCAGAT 1844  
 QY 142 GlyAlaAspProThrGlnThrSerGlnProGlnSerSerThrGlyTyrAaGlnPro 161  
 Db 1843 GGAACACAGCCACTGAGACATGATCAACCTCAATCTAGCAGGAGGCTTCAACACACCC 1784  
 QY 162 SerLeuGlyTyrGlyGlnSerAsnTyrSerTyrProGlnAlaProGlySerTyrProMet 181  
 Db 1783 AGCTTAGATATGACAGAGTACTACATGATTATCCAGGATCTGAGAGCTACCCATG 1724  
 QY 182 GlnProValThrAlaProProSerTyrProProThrSerTyrSerSerThrGlnProThr 201  
 Db 1723 CAGCAGTCACTGACCTCCATCTTACCTTACAGCTATTCCTTACAGACCGACT 1664  
 QY 202 SerTyrAspGlnSerSerTyrSerGlnGlnAsnThrTyrGlyGlnProSerSerTyrGly 221  
 Db 1663 AGTATGATCAGACAGTACTCTCAGCAGAACATATGAGGCAACGAGCAGCTATGGA 1604  
 QY 222 GlnGlnSerSerTyrGlyGlnGlnSerSerTyrGlyGlnGlnProProThrSerTyrPro 241  
 Db 1603 CAGCAGTCACTGATGATCAACAGACCTATGAGGAGGAGCTCCCATGATTACCA 1544  
 QY 242 ProGlnThrGlySerTyrSerGlnAlaProSerGlnTyrSerGlnGlnSerSerTyr 261  
 Db 1543 CCCCAACTGGATCTCAAGCAGCAGCTCAAGTCAATATGACCAACAGAGAGCTAC 1484  
 QY 262 GlyGlnGlnSerSerPheAspGlnAspHiProSerSerMetGlyValTyrGlyGlnGlu 281  
 Db 1483 GGGGAGAGAGTTATTCGACAGAGACACCCAGTGCATGGGTGTTTATGGGCAAGAG 1424  
 QY 282 SerGlyPheSerGlyProGlyGluAsnAspSerSerSerGlyProAspAsnAspGly 301  
 Db 1423 TCTGAGAGATTTTCCGACAGAGAGACCCAGCATGATGAGCTCTGATTAACCGGAGG 1364  
 QY 302 ArgGlyAspGlyGlyPheAspAspGlyGlyMetSerAspGlyGlyAspGlyGlyYarg 321  
 Db 1363 AGGGGAGAGAGGGGATTTGATCGTGAAGCATGAGAGAGTGGCGGGAGAGAGCGC 1304  
 QY 322 GlyGlyMetGlySerAlaGlyGluArgGlyGlyPheAsnAspProGlyGlyProMetAsp 341  
 Db 1303 GGTGGAATGGGC---GCTGAGAGAGCGAGTGGCTTCAATAGCCTGTGTGAGCCCATGAT 1247  
 QY 342 GlyGlyProAspLeuAspLeuGlyProProValAspProAspGluAspSerAspAsnSer 361  
 Db 1246 GAGGAGACCAATCTTATCTAGGCCACCTGTAGATCAATAGAGCTGTGACACAGT 1187  
 QY 362 AlaIleTyrValGlnGlyLeuAsnAspSerValThrLeuAspLeuAlaAspPhe 381  
 Db 1186 GCAATTATGTACAAAGATTAAATGACAGTGTGACTGTAGATGATCTGGAGAGCTTCTT 1127  
 QY 382 LysGlnGlyGlyValValIleMetAsnLysArgThrGlyGlnProMetIleHsIleTyr 401  
 Db 1126 AAGGAGGTGGGCTTTGTAAGATGAAACAAGAACTGGGCAACCAATGATCCATCTAC 1067  
 QY 402 LeuAspLysGlnThrGlyLysProLysGlyAspAlaThrValSerTyrGluAspProPro 421  
 Db 1066 CTGGACAAAGAAACAGAAAGCCCAAGGCGATGCCACATGCTCTATGAGACCCACCC 1007  
 QY 422 ThrAlaValAlaValGluTyrPheAspGlyLysAspPheGlnGlySerTyrLeuLys 441  
 Db 1006 ACTGCCAAGGCTGCGTGAATGGTTGTTGGGAAAGATTTCACAGGAGCAAACTTAAA 947  
 QY 442 ValSerLeuAlaArgLysProProMetAsnSerMetArgGlyGlyLeuProProArg 461  
 Db 946 GTCTCCCTTGCTGGAAAGAGCTTCATGAAACGATATGCGGGGTGTCTGACACCCCGT 887

QY 462 GlnGlyAspGlyMetProProProLeuAspGlyGlyProGlyGlyProGlyGlyProGly 481  
 Db 886 GAGGAGAGAGCATGACCAACCACTCCGTGAGGTCCAGAGAGCCAGAGGCTCTGGG 827  
 QY 482 GlyProMetGlyAspMetGlyGlyYargGlyGlyAspAspGlyGlyPheProProArgGly 501  
 Db 826 GAGCCATGAGGTGCGATGGAGGCGCGTGGAGAGATATGAGAGGCTTCCCTCAAGAGGA 767  
 QY 502 ProArgGlySerArgGlyAsnProSerGlyGlyGlyAsnValGlnHsAlaGlyAsp 521  
 Db 766 CCGGGGGTCCCGAGAGAAACCCCTCTGGAGGAGAAACCTCCAGCACCAGGTGGAGAC 707  
 QY 522 TyrGlnCysProAsnProGlyCysGlyAsnGlnAsnPheAlaThrPargThrGluCysAsn 541  
 Db 706 TGGCAGTGTCCCAATCCGGTTGTGAAACCAAACTTCGCTGGAGAACAGAGTGCAC 647  
 QY 542 GlnCysLeuAlaProLysProGlnGlyPheLeuProProProPheProProGlyGly 561  
 Db 646 CAGTGT-----GGT 638  
 QY 562 AspArgGlyAspGlyGlyProGlyGlyMetAspGlyGlyYargGlyGlyLeuMetAspArg 581  
 Db 637 GATCGTGGCAGAGTGGCCCTGTGTGCATGCGGGAGAGAGAGTGGCTTCATGATCGT 578  
 QY 582 GlyGlyProGlyGlyMetPheAspGlyGlyYargGlyGlyAspAspGlyGlyPheArgGly 601  
 Db 577 GGTGTCCCGGTGGATGTTTCAAGGTGGCCGTGTGGAGACAGAGTGGCTTCGTGTGT 518  
 QY 602 GlyArgGlyMetAspArgGlyGlyPheGlyGlyGlyYargArgGlyGlyProGlyGlyPro 621  
 Db 517 GCGCGGGGACATGACCGAGGTGGCTTTGGTGGAGAAAGAGAGTGGCCCTGGGGGGCCC 458  
 QY 622 ProGlyProLeuMetGlnMetGlyGlyYargArgGlyGlyYargGlyGlyProGlyGly 641  
 Db 457 CCGTGAACCTTGTATGGAAACAGATGGAGAGAAAGAGAGAGTGGAGGACCTGGAAAA 398  
 QY 642 MetAspLysGlyGlyLysArgGlnGluArgAspArgProTyr 656  
 Db 397 ATGATTAAGGCGAGCACCTCTCAGAGCCGAGAGATCGGCCCTAC 353

RESULT 7  
 US-10-094-749-1074  
 ; Sequence 1074, Application US/10094749  
 ; Publication No. US20030219741A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: ISOGAI, TAKAO  
 ; APPLICANT: SUGIYAMA, TOMOYASU  
 ; APPLICANT: OTSUKI, TETSUJI  
 ; APPLICANT: WAKAMATSU, AI  
 ; APPLICANT: SATO, HIROYUKI  
 ; APPLICANT: ISHII, SHIZUKO  
 ; APPLICANT: YAMAMOTO, JUN-ICHI  
 ; APPLICANT: ISONO, YUUKO  
 ; APPLICANT: HIO, YURI  
 ; APPLICANT: OTSUKA, KAORU  
 ; APPLICANT: NAGAI, KEIICHI  
 ; APPLICANT: IRIE, RYOTARO  
 ; APPLICANT: TAMECHIKA, ICHIRO  
 ; APPLICANT: SEKI, NAOHIKO  
 ; APPLICANT: YOSHIKAWA, TSUTOMU  
 ; APPLICANT: OTSUKA, MOTOYUKI  
 ; APPLICANT: NAGAHARI, KENJI  
 ; APPLICANT: MASUHO, YASUHIKO  
 ; TITLE OF INVENTION: NOVEL FULL-LENGTH cDNA  
 ; FILE REFERENCE: 084335/0160  
 ; CURRENT APPLICATION NUMBER: US/10/094, 749  
 ; CURRENT FILING DATE: 2002-03-12  
 ; PRIOR APPLICATION NUMBER: 60/350, 435  
 ; PRIOR FILING DATE: 2002-01-24  
 ; PRIOR APPLICATION NUMBER: JP 2001-328381  
 ; PRIOR FILING DATE: 2001-09-14  
 ; NUMBER OF SEQ ID NOS: 3381  
 ; SOFTWARE: PatentIn Ver. 2.1



SEQ ID NO 1074  
 LENGTH: 1988  
 TYPE: DNA  
 ORGANISM: Homo sapiens  
 US-10-094-749-1074

## Alignment Scores:

Pred. No.:	1.55e-227	Length:	1988
Score:	3277.00	Matches:	599
Percent Similarity:	91.31%	Conservative:	0
Best Local Similarity:	91.31%	Mismatches:	1
Query Match:	90.20%	Indels:	56
		Gaps:	1

US-10-791-017a-2 (1-656) x US-10-094-749-1074 (1-1988)

```

Qy      1 MetAlaSerThrAspTyrSerThrTyrSerGlnAlaAlaAlaGlnGlnGlyTyrSerAla 20
Db      12 ATGGCGTCCACGGATTACAGTACTTATGACCAAGCTGACGCGACGGGCTACAGTCT 71
Qy      21 TyrThrAlaGlnProThrGlnGlyTyrAlaGlnThrThrGlnAlaTyrGlyGlnGlnSer 40
Db      72 TACACGCGCCACGACCTCAAGAGTATGACACAGCCACCGGCAATGGGCAACAAAGC 131
Qy      41 TyrGlyThrTyrGlyGlnProThrAspValSerTyrThrGlnAlaGlnThrThrAlaThr 60
Db      132 TATGGAACCTATGACAGCCCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 191
Qy      61 TyrGlyGlnThrAlaTyrAlaThrSerTyrGlyGlnProProThrGlyTyrThrThrPro 80
Db      192 TATGGCAGACCCCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 251
Qy      81 ThrAlaProGlnAlaTyrSerGlnProValGlnGlyTyrGlyTyrGlyAlaTyrAspThr 100
Db      252 ACTGCCCCCAGGACATACAGCCAGCTGTCCAGGGGATAGGCACTGGTCTTATGATACC 311
Qy      101 ThrThrAlaThrAlaThrThrThrThrGlnAlaSerTyrAlaAlaGlnSerAlaTyrGlyThr 120
Db      312 ACCACGTGCTACAGTACACACACACACACACACACACACACACACACACACACACACACAC 371
Qy      121 GlnProAlaTyrProAlaTyrGlyGlnGlnProAlaAlaAlaThrAlaProThrArgProGln 140
Db      372 CAGCGCTTATTCAGCTTATGAGGACGACGACGACGACGACGACGACGACGACGACGACGAC 416
Qy      141 AspGlyAsnLysProThrGlnThrSerGlnProGlnSerSerThrGlyTyrAsnGln 160
Db      416 ----- 416
Qy      161 ProSerLeuGlyTyrGlyGlnSerAsnTyrSerTyrProGlnValProGlySerTyrPro 180
Db      416 ----- 416
Qy      181 MetGlnProValThrAlaProProSerTyrProProThrSerTyrSerSerThrGlnPro 200
Db      417 -----CTTACAGCTATTCCTTACACAGCGG 443
Qy      201 ThrSerTyrAspGlnSerSerTyrSerGlnGlnAlaAsnThrTyrGlyGlnProSerSerTyr 220
Db      444 ACTAGTTATGATCAGAGCAGTACTCTCAGCAGAACCTATAGGGCAACCGAGCAGAGTAT 503
Qy      221 GlyGlnGlnSerSerTyrGlyGlnGlnSerSerTyrGlyGlnGlnProProThrSerTyr 240
Db      504 GAGCAGACAGGGTATGATGATCAACAAAGCAGCTATAGGGCAGACGCTCCCTACATGATAC 563
Qy      241 ProProGlnThrGlySerTyrSerGlnAlaProSerGlnTyrSerGlnGlnSerSerSer 260
Db      564 CCACCCCAAACTGATCTTACAGCCAAAGCTCCAAAGTCAATATGACCAACAGAGCAGAGC 623
Qy      261 TyrGlyGlnGlnSerSerPheArgGlnAspPheProSerSerMetGlyAlaTyrGlyGln 280
Db      624 TACGGCAGAGAGATTCAATCCGACAGGACCAACCCGATAGCATTGGGTGTTATAGGGCAG 683
Qy      281 GluSerGlyGlyPheSerGlyProGlyGlnAsnArgSerMetSerGlyProAspAsnArg 300

```

```

Db      684 GAGCTCGAGAGATTTTCGACCAAGAGAACCGAGACATGATGACCTTGATTAACCGG 743
Qy      301 GlyArgGlyYArgGlyGlyPheAspArgGlyGlyMetSerArgGlyYArgGlyGlyGly 320
Db      744 GGCAGGGGAAGAGGGGGATTGATCGTGGAGCATGACAGAGGTGGCGGGGAGAGGA 803
Qy      321 ArgGlyGlyMetGlySerAlaGlyGlyAlaArgGlyGlyPheAsnLysProGlyGlyProMet 340
Db      804 CGCGGTGATATGGGACGCGCTGAGAGCGAGGTGGCTTCAATAGCTGCTGGGACCATG 863
Qy      341 AspGlyGlyProAspLeuAspLeuGlyProProValAspProAspGlyAspSerAspAsn 360
Db      864 GATGAAGACCAATCTTATGATCTTAAAGCCCTGTGATCAATGATGAAGCTTGACAC 923
Qy      361 SerAlaIleTyrValGlnGlyLeuAsnAspSerValThrLeuAspLeuAlaAspPhe 380
Db      924 AGTGCATTTTATGATACAGATTTAATGACAGTGTGATCTTATGATGATCTGGCAGACTTC 983
Qy      381 PheLysGlnCysGlyValValLysMetAsnLysArgThrGlyGlnProMetIleHisIle 400
Db      984 TTTAAGCAGTGGGGTGTGTTAAGATGACAAAGAACTGGCAACCATGATCCATC 1043
Qy      401 TyrLeuAspLysGlyLysProLysGlyAspAlaThrValSerTyrGlyLysAspPro 420
Db      1044 TACCTGACAAAGAAACAGAAAGCCAAAGGAGTCCACAGTGTCTTATGAAAGACCA 1103
Qy      421 ProThrAlaLysAlaAlaValGluTyrPheAspGlyLysAspPheGlnGlySerLysLeu 440
Db      1104 CCACCTGCCAAGGCTCCGTGGAATGTTGATGGGAATTTTCAAGGAGCAACACTT 1163
Qy      441 LysValSerLeuAlaArgLysLysProProMetAsnSerMetArgGlyLysLeuProPro 460
Db      1164 AAAGTCTCCCTGCTCGGAAAGACCTCCATGAACAGTATGCGGGGTGCTGCGCACCC 1223
Qy      461 ArgGlyGlyYArgGlyMetProProProLeuArgGlyGlyProGlyGlyYArgGlyPro 480
Db      1224 CGTGAAGGACAGAGCAGTCCACCACTCCGTGAGGTCCAGAGGCCAGGAGGTCT 1283
Qy      481 GlyGlyProMetGlyYArgMetGlyGlyYArgGlyGlyYAspArgGlyGlyPheProProArg 500
Db      1284 GGGGAGACCATGGGTGGCATGGAGGCCGTGAGAGATGAGAGAGGCTTCCCTCCAGA 1343
Qy      501 GlyProArgGlySerArgGlyYAsnProSerGlyGlyYAsnValGlnHisArgAlaGly 520
Db      1344 GAGACCCGGGGTCCGAGAGGAACCCCTCGAGAGAGAAACGTCCAGCAGCAGCTGGA 1403
Qy      521 AspTyrGlnCysProAsnProGlyYArgYAsnGlnAsnPheAlaTTPAATGThrGlyCys 540
Db      1404 GACTGGCAGTGTCCCAATCCGGGTGTGGAACCAAGAACTTCCCTGGAGAAACAGAGTGC 1463
Qy      541 AsnGlnCysLeuAlaProLysProGlyGlyPheLeuProProProPheProProProGly 560
Db      1464 AACCACTGTAAGGCCCAAGAGCTGAAGGCTTCTCCGCCACCTTTCCGCCCCGGGT 1523
Qy      561 GlyAspArgGlyYArgGlyGlyProGlyGlyMetArgGlyGlyYArgGlyGlyLeuMetAsp 580
Db      1524 GGTGATCGTGGCAGAGTGGCTGTGGTGCATCGGGGAGAGAGAGTGGCTCATAGAT 1583
Qy      581 ArgGlyGlyProGlyGlyMetPheArgGlyGlyYArgGlyGlyYAspArgGlyGlyPheArg 600
Db      1584 CGTGTGTCTCCGTGGAATGTTCAAGGTGGCGGTGTGGAGAACAGAGGTGCTTCCGT 1643
Qy      601 GlyYArgGlyYMetAspArgGlyGlyPheGlyGlyYArgArgGlyGlyProGlyGly 620
Db      1644 GGTGGCGGGGATGACCAAGTGGCTTTGTGGAGAGAAAGAGAGTGGCTTGGGGGG 1703
Qy      621 ProProGlyProLeuMetGlnMetGlyYArgArgGlyGlyYArgGlyGlyProGly 640
Db      1704 CCCCTGGAACCTTATGTAAGAACAGATGGAGAGAAAGAGAGACGTGAGAGACCTGGA 1763
Qy      641 LysMetAspLysGlyLysHisArgGlnGlyYArgArgAspArgProTyr 656

```

Db 1764 AATAATGATTAAGGCGAGACCGTCAGAGCGCAGAGATCGCCCTAC 1811

RESULT 8

US-10-439-703-58

Sequence 58, Application US/10439703

Publication No. US20040018527A1

GENERAL INFORMATION:

APPLICANT: Chang, Jenny

APPLICANT: O'Connell, Peter

TITLE OF INVENTION: Differential Patterns of Gene Expression that Predict for Docetax

TITLE OF INVENTION: Chemoresistivity and Chemoresistance

FILE REFERENCE: HO-P2482US1/10205813

CURRENT APPLICATION NUMBER: US/10/439, 703

CURRENT FILING DATE: 2003-05-16

PRIOR APPLICATION NUMBER: US 60/381,141

PRIOR FILING DATE: 2002-05-17

NUMBER OF SEQ ID NOS: 91

SOFTWARE: PatentIn version 3.1

SEQ ID NO 58

LENGTH: 1824

TYPE: DNA

ORGANISM: Human

US-10-439-703-58

Alignment Scores:

Pred. No.:	1,52e-83	Length:	1824
Score:	1282.00	Matches:	301
Percent Similarity:	53.81%	Conservative:	59
Best Local Similarity:	44.99%	Mismatches:	153
Query Match:	35.29%	Indels:	156
	17	Gaps:	30

US-10-791-017a-2 (1-656) x US-10-439-703-58 (1-1824)

QY 1 MetAlaSerThrAspTyrSerThrTyrSerGlnAlaAlaGlnGlnGlyTyrSerAla 20

Db 79 ATGGCCCTCAACAACATATATACC-----CAACACCAACCAACCAACCTATAGGGGCC 126

QY 21 TyrThrAlaGlnProThrGlnGlyTyrAla---GlnThrThrGlnAlaTyrGlyGlnGln 39

Db 127 TACCCACCAACCGCCGGGCGGCTATATCCAGACGACGACGACGCTTACGACAGCAG 186

QY 40 SerTyrGlyThrTyrGlnProThrAspValSerTyrThrGlnAlaGlnThrThrAla 59

Db 187 AGTTACAGTGCTTATAGCCAGTCACAGC-----ACTTCA 222

QY 60 ThrTyrGlnThrAlaTyrAlaThrSerTyrGlyGlnProPro---ThrGlyTyrThr 78

Db 223 GGCTATAGGCGCAGAGCTAT---TCTTCTTATGGCCAGCAGCAGACAGGCTATAGA 279

QY 79 ThrProThrAlaProGlnAlaTyrSerGlnProValGlnGlyTyrGly---ThrGlyAla 97

Db 280 ACTCAGTCACTCC-----CAGGATATAGCTCTGACTGGGGC 318

QY 98 TyrAspThrThrThrAlaThrValThrThrThrGlnAlaSerTyrThrAlaAlaGlnSerAla 117

Db 319 TATGGC-----AGTAGCCAGAGCTCCCATGCTCT 348

QY 118 TyrGlyThrGlnProAlaTyrProAlaTyrGlyGlnGlnProAlaAlaThrAlaProThr 137

Db 349 TAGGGCAGAGAGCTCTTACCTGCTATAGCCAGACGACCT----- 393

QY 138 ArgProGlnAspGlyAsnLysProThrGluThrSerGlnProGlnSerSerThrGlyGly 157

Db 394 -----CCAGCAGACCTCGGGAAGT 414

QY 158 TyrAsnGlnProSerLeuGlyTyrGlyGlnSerAsnTyrSerTyrProGlnValProGly 177

Db 415 TAGCGTAGAGCTCT-----CAGACAGACGATGCGAGCCCGC---AGTGG 462

QY 178 SerTyrProMetGlnProValThrAlaProProSerTyrProProThrSerTyrSerSer 197

Db 463 AGCTACAGCCAGACGCT-----AGCTATGTGGA 492

QY 198 ThrGlnProThrSerTyrAspGlnSerSerTyrSerGlnGlnAlaThrTyrGlyGlnPro 217

Db 493 -----CAGCAGAAAGCTATGACAGCAGAAAGCTATTAATCCCTT 534

QY 218 SerSerTyrGlyGlnGlnSerSerTyrGlyGlnGlnSerSerTyrGlyGlnGlnProPro 237

Db 535 CAGGCTATGACAGCAGAAACGATACACAGACAGATGCTGTGAGGTGAGAGTGA 594

QY 238 ThrSerTyrProProGlnThrGlySerTyrSerGlnAlaProSerGlnTyrSerGlnGln 257

Db 595 GGTGCA-----GGTACTATGCGCAAGATCAATCTTCATGATGATGCT 639

QY 258 SerSerSerTyrGlyGlnGlnSerSerPheAspGlnAspHisProSerSerMetGlyVal 277

Db 640 GGTGGCAGGTGCGCGCTTATGCAATCAAGACCAAGGTGTGAGGTGACCGGTGCG 699

QY 278 TyrGlyGlnGlnSerGlyGlyPheSerGlyProGlyGlyLysAspSerMetSerGlyPro 297

Db 700 TATGACAGACGACCGTGA-----GCTCGCGCAGCGGTGACATGCT--- 744

QY 298 AspAsnArgGlyLysGlyLysGlyGlyPheAspArg-----GlyGlyMetSerArgGly 315

Db 745 GCGCGCGCGCGCGCGCGGTGCTTCAACCGCAGCAGTGTGCTATGAACCCAGA 804

QY 316 GlyArgGlyGlyLysLysArgGlyGlyMetGlySerAlaGly-----GlyArgGlyGlyPhe 333

Db 805 GGTGTGTGAGGTGCGCGGTGAGGACAGAGGTGCGATGCGCGAGTACCTGTGTGCTTC 864

QY 334 AsnLysProGlyGlyLysProMetAspGlnGlyProAspLeuAspLeuGlyProProValAsp 353

Db 865 AATTAATTTGTGTGCGCTCGGACCAAGATACGTCATGAC-----TCC 909

QY 354 ProAspGlnAspSerAspAsnSerAlaIleTyrValGlnGlyLeuAsnAspSerValThr 373

Db 910 GAACAGAGATATATCAACAACAACATCTTGTCCAGGCTGTGCGAATGTTACA 969

QY 374 LeuAspAspLeuAlaAspPheLysGlnGlyValValLysMetAsnLysAspArgThr 393

Db 970 ATTGAGTCTGTGCTATTAATTTGTTACAGAGATTTGTTATTAAGCAACAAGAAACG 1029

QY 394 GlyGlnProMetGlnIleHisIleTyrLeuAspLysGlnThrGlyLysProLysGlyAspAla 413

Db 1030 GACACGCCCATGATTAATTTGTTACAGAGAGAAATCGCAAGCTGAAGGAGAGCA 1089

QY 414 ThrValSerTyrGlnAspProProThrAlaLysAlaAlaValGluTyrPheAspGlyLys 433

Db 1090 ACGTCTCTTTGATGATACCACTTCAGCTAAAGCAGCTAATGACGTGTTGATGTAA 1149

QY 434 AspPheGlnGlySerLysLeuLysValSerLeuAlaArgLysLysProProMetAsnSer 453

Db 1150 GAATTTCCGAATCTCTTCAAGGTCTCATTTGCTACTGCGCGGACAGCTTAAT--- 1206

QY 454 MetArgGlyLysLeuProProArgGlnGlyArgGlyMetProProProLeuAspArgGly 473

Db 1207 ---CGGGGTGTGGCAAGTGTGTGAGGCCAGG-----CGAGGAGCA 1248

QY 474 ProGlyGlyProGlyGlyProGlyGlyProMetGlyArgMetGlyLysArgGlyAsp 493

Db 1249 CCGATGGCGGTGAGAGGCTATGAGGTGT-----GGCAGTGTGTGTGTGCG 1296

QY 494 ArgGlyLysPheProProArgGlyProArgGlySerArgGlyAsnProSerGlyGlyGly 513

Db 1297 CAGAGAGATTTCCAGTGAAGT-----GGTGGCGGT 1329

QY 514 AsnValGlnHisArgAlaAspTyrGlnCysProAsnProGlyCysGlyAsnGlnAsn 533

Db 1330 GAGAGACAGACAGAGCTGTGATCTGAAGTGTCCATATCCACCTGTGATATGAAC 1389

QY 534 PheAlaTyrArgThrGlnCysAsnGlnCysLysAlaLysProLysProGlnGlyPheLeuPro 553

Db 1390 TTCTCTTGAGAGAAATGAATGACAACAGTGTAAAGCCCTTAAACCAATATGCG----- 1440

```

QY 554 ProProheProProGlyYAspArgGlyYArgGlyYProGlyYGlyMetArg--- 572
DB 1441 -----CAGGA-----GGGGACAGGTGGCTCTCACATG 1470
QY 573 GlyYArgGlyYLeuMetAspArgGlyYProGlyYMetPheArgGly----- 590
DB 1471 GGGGGTAACAGAGGGATGATCGTCGTGGAGAGAGGCTATATGACGGCGGTAC 1530
QY 591 ---GlyYArgGlyYAspArgGlyYArgGlyYArgGlyYArgGlyYArgGlyY 608
DB 1531 CGGGGGCCGGCGGGAGCCGTGAGAGGCTCCAGGGGGCGGGGTGGGACAGAGT 1590
QY 609 GlyPheGlyYArgGlyYArgGlyYArgGlyYProGlyYProProGlyYProLeuMetGln 628
DB 1591 GGGCTT-----1596
QY 629 MetGlyYArgGlyYArgGlyYArgGlyYProGlyYMetAsp---LysGlyYGlnHis 647
DB 1597 -----GGCCTGGCAAGATGATTCACAGGGGTGAGCAC 1629
QY 648 ArgGlnGlnArgArgAspArgProTyr 656
DB 1630 AGACAGGATCGCAGGAGGAGGCGGTAT 1656

```

## RESULT 9

```

US-10-755-889-649
/ Sequence 649, Application US/10755889
/ Publication No. US20040171823A1
/ GENERAL INFORMATION:
/ APPLICANT: Bristol-Myers Squibb Company
/ TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES ASSOCIATED WITH THE NF-KB
/ FILE REFERENCE: D0284 NP
/ CURRENT APPLICATION NUMBER: US/10/755, 889
/ PRIOR FILING DATE: 2004-01-13
/ PRIOR APPLICATION NUMBER: U.S. 60/440, 068
/ PRIOR FILING DATE: 2003-01-14
/ PRIOR APPLICATION NUMBER: U.S. 60/469, 757
/ NUMBER OF SEQ ID NOS: 823
/ SOFTWARE: Patent version 3.2
/ SEQ ID NO 649
/ LENGTH: 1822
/ TYPE: DNA
/ ORGANISM: Homo sapiens
US-10-755-889-649

```

## Alignment Scores:

```

Pred. No.: 1 65e-83 Length: 1822
Score: 1281.50 Matches: 299
Percent Similarity: 53.59% Conservative: 59
Best Local Similarity: 44.76% Mismatches: 155
Query Match: 35.27% Indels: 155
DB: 18 Gaps: 29

```

US-10-791-017A-2 (1-656) x US-10-755-889-649 (1-1822)

```

QY 1 MetAlaSerThrAspTyrSerThrTyrSerGlnAlaAlaGlnGlnGlyTyrSerAla 20
DB 73 ATGGCTCAACAGATTATACC-----CAACAACAACCAAGACTATGGGGCC 120
QY 21 TyrThrAlaGlnProThrGlnGlyTyrAla---GlnThrThrGlnAlaTyrGlyGln 39
DB 121 TACCCACACCCAGCCCGGCGGCTATTTCCACAGAGCGAGTCCGCTACGACAGCAG 180
QY 40 SerTyrGlyThrTyrGlyGlnProThrAspValSerTyrThrGlnAlaGlnThrAla 59
DB 181 AGTTACAGTGTATAGCCAGTCCAGAC-----ACTTCA 216
QY 60 ThrTyrGlyGlnThrAlaTyrAlaThrSerTyrGlyGlnProThrGlyTyrThrThr 79
DB 217 GGCTATGGCCAGAGCAGCTAT---TCTTCTTATGGCCAGAGCAGAAACAGCTATGAACT 273

```

```

QY 80 ProThrAlaProGlnAlaTyrSerGlnProValGlnGlyTyrGly---ThrGlyAlaTyr 98
DB 274 CAGTCAACTCC-----CAGGATATGGCTCCAGCTGGCGGTAT 312
QY 99 AspThrThrThrAlaThrValThrThrThrGlnAlaSerTyrAlaAlaGlnSerAlaTyr 118
DB 313 GGC-----AGTACCAAGAGCTCCAACTCTTAC 342
QY 119 GlyThrGlnProAlaTyrProAlaTyrGlyGlnGlnProAlaAlaThrAlaProThrArg 138
DB 343 GGGCAGACAGCTCTTATCTGCTATGGCAGACAGCAGCT-----384
QY 139 ProGlnAspGlyValenylsProThrGlnThrSerGlnProGlnSerThrGlyY Tyr 158
DB 385 -----CCGAGCAGCAGCTCGGAAGTTAC 408
QY 159 AsnGlnProSerLeuGlyTyrGlyGlnSerAsnTyrSerTyrProGlnValProGlySer 178
DB 409 GGTAGCAGTTCT-----CAGACAGCAGCTATGGCAGCCCCAG---AGTGGAGC 456
QY 179 TyrProMetGlnProValThrAlaProProSerTyrProProThrSerTyrSerSerThr 198
DB 457 TACAGCCAGCAGCT-----AGCTATGTGGA---483
QY 199 GlnProThrSerTyrAspGlnSerSerTyrSerGlnGlnAsnThrTyrGlyGlnProSer 218
DB 484 -----CAGCAGCAAGCTATGACACAGCAAGCTATATATCCCTCCAG 528
QY 219 SerTyrGlyGlnGlnSerSerTyrGlyGlnGlnSerSerTyrGlyGlnGlnProProThr 238
DB 529 GGCTATGACAGCAAGCAAGCTATCAACAGCAGAGTGTGTGAGTGTGAGTGTGAGTGT 588
QY 239 SerTyrProProGlnThrGlySerTyrSerGlnAlaProSerGlnTyrSerGlnGlnSer 258
DB 589 GGA-----GGTACTATGGCCAGATCAATCTCCATGAGTGTGTGTGTGTGTGTGT 633
QY 259 SerSerTyrGlyGlnGlnSerSerPheArgGlnAspHisProSerSerMetGlyValTyr 278
DB 634 GGCAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 693
QY 279 GlyGlnGlnSerGlyYArgGlyYArgGlyYProGlyYGlnAsnArgSerMetSerGlyProAsp 298
DB 694 GAGCAGCAGAGCCGTGA-----GGCCGGCGCAGGGGTGTGCACTGTGT---GCC 738
QY 299 AsnArgGlyYArgGlyYArgGlyYArgGlyYArgGlyYArgGlyYArgGlyYArgGlyY 316
DB 739 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 798
QY 317 ArgGlyYArgGlyYArgGlyYArgGlyYArgGlyYArgGlyYArgGlyYArgGlyYArgGlyY 334
DB 799 CGTGGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 858
QY 335 LysProGlyYArgProMetAspGlnGlyProAspLeuAspLeuGlyProProValAspPro 354
DB 859 AATTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 903
QY 355 AspGluAspSerAspAsnSerAlaIleTyrValGlnGlyLeuAsnAspSerValThrLeu 374
DB 904 CAGGATATTCAGACAACAACACATCTTTGTGCAAGGGCTGTGTGAGAAATGTATCAATT 963
QY 375 AspAspLeuAlaAspPhePheLeuGlnCyGlyValValIleMetLeuAsnTyrGly 394
DB 964 GAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1023
QY 395 GlnProMetIleHisIleTyrLeuAspLysGlnThrGlyYAspProLysGlyAspAlaThr 414
DB 1024 CAGCCATATATTTATTTGTACACAGCAGAGAACTGTGCACTGAGAGGAGGAGCAG 1083
QY 415 ValSerTyrGluAspProProThrAlaValAlaAlaValGlnThrPheAspGlyAsp 434
DB 1084 GTCTCTTTGATGATCACTTCACTTAAAGCAGCTATGACTGTGTGTGTGTGTGTGTGT 1143
QY 435 PheGlnGlySerIleValLeuValSerLeuAlaArgIleValLeuProProMetAsnSerMet 454

```



```

Db      807 GGTCTGTGAGAGTGGCCCTGAGACGAGGTGGCATGGGCGGAAATGACCTGTGGTTC 866
QY      334 AenlyseProglYglYProMetAapglYglYProApleuApleuGlYProProValAap 333
Db      867 AATATAATTTGGTGGCCCTCGGAGACCAAGATGATCATCGTATGAC-----TCC 911
QY      354 ProAapglAapSerAapAapSerAalIleTyValGlnIleuAapSerValThr 373
Db      912 GAACAGATATATTCAGACCAACCACTTTGTGTGAAGCCCTGGGTGGAATGTATCA 971
QY      374 LeuAapAapLeuAalAapPhePheYglYcysGlyValValIlyseAanlyseArgThr 393
Db      972 ATTGAGCTGTGCTGATTTACTTCAAGCAGATTGGTTATTAAAGAAACCAAGAAACG 1031
QY      394 GlyGlnProMetIleHsiIleTyLeuAapIlyseGlnThrIlyYseProIlyseIlyAapAla 413
Db      1032 GGAAGAGCCCATGATTATTTGTACACAGACAGGAAACCTGGCAAGCTGAAGGAGAGCA 1091
QY      414 ThrValSerTyglYleuAapProProThraIleAlaIleAlaIleGlnTTPheAapGlyLys 433
Db      1092 ACGGTCTCTTTATGACCACTTCAGCTAAAGCAGCTTATGACGTGTGATGTAA 1151
QY      434 AapPheGlnIlyseTyLeuIlyseValSerLeuAalAargIlyIlyseProMetAapSer 453
Db      1152 GAATTCCTCCGAAATCTATCAAGGTCTACTTGTCTACTCTCGCGGAGACTTTAAT--- 1208
QY      454 MetAargGlyglYleuProProAargGlnIlyArgGlyMetProProProleuAargIly 473
Db      1209 ---CGGGGTGTGTGCAATGTGCTGAGAGCCGAGGG-----CGAGAGAGA 1250
QY      474 ProglYglYProglYglYProglYglYProMetGlyIlyArgGlyIlyAap 493
Db      1251 CCATAGGCGCGTGAAGCTATGAGAGTGTG-----GGCAGTGTGTGTGTGGGC 1298
QY      494 AargGlyglYPheProProAargGlyProAargGlySerAargIlyAapProSerGlyglY 513
Db      1299 CAGAGAGGATTTCCCACTGAGGT-----GGTGGCGGT 1331
QY      514 AenValGlnIleAargAlaGlyAapTTPGlnCyseProAapProglYcysGlyAanGlnAap 533
Db      1332 GAGAGACAGACAGAGCTGTGTGACTGGAAGTCTCTAATCCCACTGTGAATATATGAAAC 1391
QY      534 PheAlaIleThraArgThrGlnCyseAanGlnCyseIlyAalAapProIlyProglYglYPheLeuPro 553
Db      1392 TTCTCTTGAAGATGATCAACCAAGTGAAGGCCCTTAACCAAGATGGC----- 1442
QY      554 ProProPheProProProglYglYAapAargGlyIlyArgGlyglYProglYglYMetAarg--- 572
Db      1443 -----CCAGGA-----GGGGGACCAAGGTGGCTCTCACATG 1472
QY      573 GlyIlyAargGlyglYleuMetAapAargGlyglYProglYglYMetPheAargIly----- 590
Db      1473 GGGGGTAACTACAGGGGATGATCGTGTGTGCGACAGAGGCTATATGATCAGAGGGGTAC 1532
QY      591 ---GlyAargGlyglYleuAapAargGlyglYPheAargGlyglYIlyArg---GlyMetAapAargGly 608
Db      1533 CGGGGGCGCGCGGAGACCGTGTGAGGCTTCCAGGGGGCGGGGTGTGGGACAGAGGT 1592
QY      609 GlyPheGlyglYglYIlyAargAargGlyglYProglYglYProProglYProleuMetGlnGln 628
Db      1593 GGGCTTT-----GGCCCTGGCAAGATGATTCACAGGGGTAGACAC 1631
QY      629 MetGlyIlyAargAargGlyglYIlyArgGlyglYProglYglYMetAap---IlyGlyglYIleHis 647
Db      1599 -----GGCCCTGGCAAGATGATTCACAGGGGTAGACAC 1631
QY      648 ArgGlnGlnAargAargAapArgProTy 656
Db      1632 AGACAGATGACGAGGAGGCGGTAT 1658

```

RESULT 11  
 US-10-425-115-182496  
 ; Sequence 182496, Application US/10425115

```

; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovacic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCES: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; NUMBER OF SEQ. ID NOS: 369326
; SEQ. ID NO. 182496
; LENGTH: 763
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(763)
; OTHER INFORMATION: unsure at all n locations
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_98018C.1
; US-10-425-115-182496

Alignment Scores:
Pred. No.: 2,43e-79 Length: 763
Score: 1219.00 Matches: 225
Percent Similarity: 97.86% Conservative: 4
Best Local Similarity: 96.15% Mismatches: 5
Query Match: 33.55% Indels: 0
DB: 18 Gaps: 0

US-10-791-017A-2 (1-656) x US-10-425-115-182496 (1-763)
QY      1 MetAlseerThraApyrSerThryrSerGlnAlaAlaIleGlnGlnIlyrSerAla 20
Db      60 ATGGCTCCACGATTTACATTAAGTCAAGCTGACGCCCAAGGCTACAGTGTCT 119
QY      21 TyrThraIleGlnProThrglnIlyTyraIleGlnThrGlnAlaIlyrGlyGlnIleSer 40
Db      120 TACACGCCCGCAGCACTCAAGATATGACAGACCCAGGCAATATGGGCAACAAAGC 179
QY      41 TyrglyThryrtyglYglYleuProThraApyrSerTyrrThrglnAlaGlnThrrThraIleThr 60
Db      180 TATGAACCTATGAGACAGCTTACTGATGTCACTAATACAGCTCAAGACCTGACACC 239
QY      61 TyrglyGlnThraIleThraIleThraIleThraIleThraIleThraIleThraIleThraIle 80
Db      240 TACGGGCACTCAATATGATTAAGTCTTACGAGAGCTTCCACTGTATATGACATCCA 299
QY      81 ThraIleProGlnAlaTyrrSerGlnProValGlnIlyTyrglyThryrtyglYIlyAapThr 100
Db      300 ACTGCCCGGAGCGTACAGACGCTGTGCAAGGCTATATGAGCACTTATATGACAGC 359
QY      101 ThrrThraIleThrrThrrThrrThrrThrrThrrThrrThrrThrrThrrThrrThrrThrr 120
Db      360 ACCACGTCTCAAGTCAACCAAGAGGCTCTTACGCAAGCTCAAGCTCAAGTATATGAGACC 419
QY      121 GlnProAlaTyrrProAlaTyrrGlnIleGlnProAlaAlaThrrAlaProThrrAapProGln 140
Db      420 CAGCTGTCTTACCCCACTATGAGCAGACCAAGCAAGCCAGCCCACTTACCAAGCCAG 479
QY      141 AapGlyAapIlyAapProThrglnThrrSerGlnProGlnIleSerSerThryrtyglYTyraGln 160
Db      480 GATGTAAACAGCCTGTGAGCTAATGACACTCAATCTTACAGAGGGGTATTAACAA 539
QY      161 ProSerIleuGlnIlyrtyglYleuSerAenTyrrSerTyrrProGlnIlyProglYserTyrrPro 180
Db      540 CCACAGCTAGATATGAGACAGATTAATCAAGCTATCCCAAGATACCTTGGAGCTACCA 599
QY      181 MetGlnProValThrrAlaProProSerTyrrProProThrrSerTyrrSerSerThrrGlnPro 200
Db      600 ATGCAAGCACTCAAGCAAGCTCTCAATCTTATCTTACCAAGCTATCTCTTCAAGCCG 659

```

Qy 201 ThSerTyraapginserSerTyrsSerGlnGlnaanthTyrglyglnProserSerTy 220  
Db 660 ACTAGNTACATCAAGACAGTACTCTCAAGACACCTTAGGGCCGAGCACTAT 719  
Qy 221 GlyGlnInserSerTyrglyglnInserSerTyrglygln 234  
Db 720 GACAGCAGAGTAGCTATGCTCAACAAAGCAGCTATGGGAG 761

RESULT 12  
US-10-198-846-9847  
; Sequence 9847, Application US/10198846  
; Publication No. US2003009974A1  
; GENERAL INFORMATION:  
; APPLICANT: Little, James  
; APPLICANT: Xu, Yongyao  
; APPLICANT: Wang, Youzhen  
; APPLICANT: Steilmann, Kathleen  
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS  
; TITLE OF INVENTION: FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND  
; TITLE OF INVENTION: THERAPY OF BREAST CANCER  
; FILE REFERENCE: MRI-049  
; CURRENT APPLICATION NUMBER: US/10/198,846  
; CURRENT FILING DATE: 2002-07-18  
; PRIOR APPLICATION NUMBER: 60/306,220  
; PRIOR FILING DATE: 2001-07-18  
; NUMBER OF SEQ ID NOS: 14084  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 9847  
; LENGTH: 2299  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-198-846-9847

Alignment Scores:  
Pred. No.: 3,67e-76 Length: 2299  
Score: 1181.00 Matches: 222  
Percent Similarity: 66.07% Conservative: 0  
Best Local Similarity: 66.07% Mismatches: 1  
Query Match: 32.51% Indels: 114  
DB: 14 Gaps: 1

US-10-791-017a-2 (1-656) x US-10-198-846-9847 (1-2299)

Qy 338 GlyProMeRaapginuglyProaapleuapleuglyProProvaJaapProaapGluap 357  
Db 1072 GGAACCCATGATGAAGACAGATCTGATCTAGGCCACCTGTAGATCCAGATGAAGAC 1131  
Qy 358 SerAspaanSerAlaIetyrvaGlnGlyLeuAnaapSerValThrLeuAapleu 377  
Db 1132 TCTGACAAAGTGAATTTATGATCAAGATTAATGACAGTGTGATCTAGATGATCTG 1191  
Qy 378 AlaAapPhehelyglnGlyValVallyllyllyllyllyllyllyllyllyllyl 397  
Db 1192 GCAGACTCTTTAGAGTGTGGGTGTGTAAGTGAAGAAAGAACTGGGCAACCATG 1251  
Qy 398 IleHisIleTyraapleuapleuglyllyllyllyllyllyllyllyllyllyl 417  
Db 1252 ATCCACATCTACTCGGACAGAAAGAAAGCCCAAGGCAAGTCAAGTATGCGGGGTGT 1311  
Qy 418 GluAapProProThralalyalaIalValGluTrPheAapGlyllyllyllyllyl 437  
Db 1312 GAAAGACCAACCACTGCGCAAGGCGCGGTGAAGTGAAGTGAAGTGAAGTGAAGG 1371  
Qy 438 SerTyraapleuapleuapleuapleuapleuapleuapleuapleuapleuap 457  
Db 1372 AGCAAACTTAAGTCT 1431  
Qy 458 LeuProProaapginuglyllyllyllyllyllyllyllyllyllyllyllyl 477  
Db 1432 CTGCGACCCCGTGTGGGCAAGGAGTCAACCACTCTCTCTCTCTCTCTCTCTCTCT 1491  
Qy 478 GlyGlyProGlyllyllyllyllyllyllyllyllyllyllyllyllyllyllyl 497

Db 1492 GAGAGCTCTGGGGGACCACTGGGTGCAATGGAGGCGGTGAGAGATAGAGAGCTTC 1551  
Qy 498 ProProaapGlyProaapGlySeratGlyllyllyllyllyllyllyllyllyllyl 517  
Db 1552 CTTCCAAAGAGAGACCCCGGGGTTCCTCCAGAGAAACCTCTCTGAGAGAGAAACGTCAGCAC 1611  
Qy 518 ArgAlaGlyAapTyraapleuapleuapleuapleuapleuapleuapleuapleu 528  
Db 1612 CGAGCTGAGACTGAGAGTGTCTCCATCTCCGTA-TGTACTTCTTGGCAATTGATACC 1670  
Qy 528 ----- 528  
Db 1671 TACAGTGAAGCAACCTTCCTCCACCCATCCCACTAGAGTGATGCTCTGTCTA 1730  
Qy 528 ----- 528  
Db 1731 GAGAAAGAAATGATGATCCCTGATGGCTGTAGGACATGATGATGATGATGATGAT 1790  
Qy 528 ----- 528  
Db 1791 GCTTCAAGCTCTTGAAGATTTGATGATGATGATGATGATGATGATGATGATGAT 1850  
Qy 528 ----- 528  
Db 1851 TGTGCTTAAGCATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1910  
Qy 528 ----- 528  
Db 1911 CACCACTTCTCTGTTATCTCTCTTATGATGATGATGATGATGATGATGATGATGAT 1970  
Qy 529 ----- 529  
Db 1971 GATGAGAGGTTGTGAAACCAAGATCTGCTGAGAAAGAGAGTCAAGCTGTAG 2030  
Qy 545 AlaProlyProGlyllyllyllyllyllyllyllyllyllyllyllyllyllyllyl 560  
Db 2031 GCCCAAGCTGAAAGCTTCTCCGCAACCTTCCGCGGAGT 2078

RESULT 13  
US-09-918-995-27690  
; Sequence 27690, Application US/09918995  
; Publication No. US20030073623A1  
; GENERAL INFORMATION:  
; APPLICANT: Hyseq, Inc.  
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED  
; FILE REFERENCE: 20411-756  
; CURRENT APPLICATION NUMBER: US/09/918,995  
; CURRENT FILING DATE: 2001-07-30  
; PRIOR APPLICATION NUMBER: US/09/235,076  
; PRIOR FILING DATE: 1999-01-20  
; NUMBER OF SEQ ID NOS: 38054  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 27690  
; LENGTH: 550  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)...(550)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-918-995-27690

Alignment Scores:  
Pred. No.: 7.63e-61 Length: 550  
Score: 961.00 Matches: 166  
Percent Similarity: 97.65% Conservative: 0  
Best Local Similarity: 97.65% Mismatches: 4  
Query Match: 26.45% Indels: 0  
DB: 10 Gaps: 0

US-10-791-017a-2 (1-656) x US-09-918-995-27690 (1-550)



QY 437 G1ySer1yLeu1yVal1Ser1Leu1AArg1yLys1yPro1yMet1yCys1ySer1Met1yArg1y 456  
DB 40 GGGAGCAAACTTAAAGTCTCTGCTGGAGAAAGCTTCAATGAACATATAGCGGTGT 99  
QY 457 G1yLeu1yPro1yArg1yLys1yArg1yMet1yPro1yPro1yLeu1yArg1yLys1yPro1yGly 476  
DB 100 GGTCTGACCACTCCGTGAGGCAAGGCAATGCCACCACTCCGTGAGGCTCAAGAGGC 159  
QY 477 Pro1yGly1yPro1yLys1yPro1yMet1yArg1yMet1yGly1yArg1yLys1yArg1yLys1y 496  
DB 160 CCAAGAGGTCTCTGAGGAGCCCATGGGTCCGATGGAGAGCCGTGAGAGGATAGAGAGGC 219  
QY 497 Phe1yPro1yArg1yLys1yPro1yArg1ySer1yArg1yLys1yPro1ySer1yGly1yVal1yGln 516  
DB 220 TTCTCTTCAAGAGAGCACTCCGGGGTTCCTGGAGGAACTCTGAGAGAGAAAGTCCAG 279  
QY 517 H1e1yArg1yAla1yAsp1yTyr1yGly1yPro1yLys1yPro1yLys1yVal1ySer1yLeu1yAla1yTyr 536  
DB 280 CACCAAGCTGAGAGCTGGAGTGTCCCAATCCGGGTGTGAAACCAAGACTTGGCTGG 339  
QY 537 Arg1yThr1yGly1yAla1yGln1yCys1yLys1yAla1yPro1yLys1yLys1yPhe1yLeu1yPro1yPhe 556  
DB 340 AGAAGAGAGTGAACCAAGCTTAAGGCCCAAGCTTGAAGCTTCCCTCCGCACTTT 359  
QY 557 Pro1yPro1yArg1yLys1yAsp1yArg1yLys1yPro1yLys1yMet1yArg1yLys1yArg1y 576  
DB 400 CCCCCCGGGTGTATCGTGGCAGAGTGGCCCTGTGTGCATGCGGAGAGAGAGT 459  
QY 577 G1yLeu1yMet1yAsp1yArg1yLys1yPro1yLys1yMet1yPhe1yArg1yLys1yArg1yLys1y 596  
DB 460 GGGCTCATGGATGTGTGTGTCCGGGTGAATGTTCAATAGTGTGCGGTGTGAGAGACA 519  
QY 597 G1yGly1yPhe1yArg1yLys1yArg1yMet1yAsp 606  
DB 520 GGTGGCTTCGTGTGTGGCCGGGGCATGTGAC 549

RESULT 14  
US-09-918-995-9557  
/ Sequence 9557, Application US/09918995  
/ Publication No. US2003073623A1  
/ GENERAL INFORMATION:  
/ APPLICANT: Hysed, Inc.  
/ TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED  
/ FILE REFERENCE: 20411-756  
/ CURRENT APPLICATION NUMBER: US/09/918,995  
/ CURRENT FILING DATE: 2001-07-30  
/ PRIOR APPLICATION NUMBER: US/09/235,076  
/ PRIOR FILING DATE: 1999-01-20  
/ NUMBER OF SEQ ID NOS: 38054  
/ SOFTWARE: FastSeq for Windows Version 3.0  
/ SEQ ID NO 9557  
/ LENGTH: 568  
/ TYPE: DNA  
/ ORGANISM: Homo sapiens  
/ FEATURE:  
/ NAME/KEY: misc.feature  
/ LOCATION: (1)...(568)  
/ OTHER INFORMATION: n = A,T,C or G  
US-09-918-995-9557

Alignment Scores:  
Pred. No.: 3,74e-54 Length: 568  
Score: 868.50 Matches: 166  
Percent Similarity: 97.08% Conservative: 0  
Best Local Similarity: 97.08% Mismatches: 4  
Query Match: 23.91% Indels: 2  
DB: 10 Gaps: 1

US-10-791-017A-2 (1-656) x US-09-918-995-9557 (1-568)

QY 307 PheAspArgGlyLysMetSerArgGlyArgGlyLysArgGlyLysMetGlySer 326

DB 56 TTGATGTGGAGGCAATGACCAAGGTGGCGGAGAGAGAGCGGTGGAAATGGCC--- 112  
QY 327 AlAG1yGly1yLys1yGly1yPhe1yLys1yPro1yGly1yPro1yMet1yArg1yLys1yPro1yLeu 346  
DB 113 GCTGAGAGAGAGGTGGCTTCAATTAAGCTGTGTGAGCCATGTGATGAAGACCAAGATCTT 172  
QY 347 Asp1yLeu1yPro1yVal1yAsp1yPro1yArg1yLys1ySer1yAsp1ySer1yAla1yTyr1yVal1yGln 366  
DB 173 GATCTGAGCCCACTGTAGATTCAGATGAAGACTTGAACAACAGTGCATTTATGTACAA 232  
QY 367 G1yLeu1yAsp1ySer1yThr1yLeu1yAsp1yLeu1yAsp1yPhe1yLys1yGly1yVal1 386  
DB 233 GGAATTAATACAGTGTGACTTGAATGATCTGACAGACTTCTTTAAGCAGTGTGGGT 292  
QY 387 Val1yMet1yAsn1yLys1yThr1yGly1yLys1yPro1yMet1yGly1yLeu1yAsp1yLys1yThr 406  
DB 293 GTTAAGATGAACAAGAGAACTGGGCAACCATGATCCATCTGACCAAGAAACA 352  
QY 407 G1yLys1yPro1yLys1yAsp1yAla1yThr1ySer1yGly1yAsp1yPro1yThr1yAla1yAla 426  
DB 353 GGAAGCCCAAGGCAATGCCACAGTGTCTTATGAAGACCACTGCAAGGCTGCC 412  
QY 427 Val1yGly1yThr1yPhe1yAsp1yLys1yAsp1yGln1ySer1yLeu1yVal1ySer1yLeu1yArg 446  
DB 413 GTGAAATGTATTATGGAGAAATTTCAAGGAGCAAACTTAAGTCTCTGCTGG 472  
QY 447 Lys1yLys1yPro1yMet1yAsn1yMet1yArg1yGly1yLeu1yPro1yArg1yLys1yArg1yMet 466  
DB 473 AAGAAGCTTCAATGAACAGTATGCC-GGTGGCTGCCACCCGTGAGGCAAGAGCATG 531  
QY 467 Pro1yPro1yLeu1yArg1yLys1yPro1yGly1yPro 477  
DB 532 CCAACCACTTTCGTGAGGTCCAGAGAGGCCCC 564

RESULT 15  
US-09-864-761-8001  
/ Sequence 8001, Application US/09864761  
/ Patent No. US20020048763A1  
/ GENERAL INFORMATION:  
/ APPLICANT: Penn, Sharon G.  
/ APPLICANT: Rank, David R.  
/ APPLICANT: Hanzel, David K.  
/ APPLICANT: Chen, Wensheng  
/ TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
/ FILE REFERENCE: Aeomica-X-1  
/ CURRENT APPLICATION NUMBER: US/09/864,761  
/ CURRENT FILING DATE: 2001-05-23  
/ PRIOR APPLICATION NUMBER: US 60/180,312  
/ PRIOR FILING DATE: 2000-02-04  
/ PRIOR APPLICATION NUMBER: US 60/207,456  
/ PRIOR FILING DATE: 2000-05-26  
/ PRIOR APPLICATION NUMBER: US 09/632,366  
/ PRIOR FILING DATE: 2000-08-03  
/ PRIOR APPLICATION NUMBER: GB 24263,6  
/ PRIOR FILING DATE: 2000-10-04  
/ PRIOR APPLICATION NUMBER: US 60/236,359  
/ PRIOR FILING DATE: 2000-09-27  
/ PRIOR APPLICATION NUMBER: PCT/US01/00666  
/ PRIOR FILING DATE: 2001-01-30  
/ PRIOR APPLICATION NUMBER: PCT/US01/00667  
/ PRIOR FILING DATE: 2001-01-30  
/ PRIOR APPLICATION NUMBER: PCT/US01/00664  
/ PRIOR FILING DATE: 2001-01-30  
/ PRIOR APPLICATION NUMBER: PCT/US01/00669  
/ PRIOR FILING DATE: 2001-01-30  
/ PRIOR APPLICATION NUMBER: PCT/US01/00665  
/ PRIOR FILING DATE: 2001-01-30  
/ PRIOR APPLICATION NUMBER: PCT/US01/00668  
/ PRIOR FILING DATE: 2001-01-30  
/ PRIOR APPLICATION NUMBER: PCT/US01/00663  
/ PRIOR FILING DATE: 2001-01-30



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

# OM protein - nucleic search, using frame\_plus.p2n model

Run on: February 20, 2005, 21:44:36, Search time 4920 Seconds  
(without alignments)  
5075.236 Million cell updates/sec

Title: US-10-791-017A-2

Perfect score: 3633  
Sequence: 1 MASTDYSTYSQAAAGQCYSA.....GGPKNDKXGHRDRDPY 656

Scoring table:  
BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 34239544 seqs, 19032343700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

## Command line parameters:

-MODE=frame+ p2n.model -DEV=xlp  
-Q/cgml\_1/USFTO.spool\_p/US10791017/runat\_17022005\_125807\_22086/app\_query.fasta\_1.1358  
-DB=EST -QPMF=fastap -SUFFIX=p2n.rst -MIMATCH=0.1 -LOOPCT=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blomsum62 -TRANS=human40.cdd -LIST=45  
-DOCALLIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=pro -NOR=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=us10791017 @CGML\_1.1 6628 @runat\_17022005\_125807\_22086 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -MAIN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

## Database :

EST: \*  
1: gb\_est1.\*  
2: gb\_est2.\*  
3: gb\_hic.\*  
4: gb\_est3.\*  
5: gb\_est4.\*  
6: gb\_est5.\*  
7: gb\_est6.\*  
8: gb\_ges1.\*  
9: gb\_ges2.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3633	100.0	2157	3	CR604952 full-leng
2	3633	100.0	2157	3	CR608375 full-leng
3	3618.5	99.6	2159	3	CR625247 full-leng
4	3615	99.5	2103	3	CR619493 full-leng
5	3600.5	99.1	2070	3	CR602561 full-leng
6	3600.5	99.1	2093	3	CR594868 full-leng
7	3600.5	99.1	2126	3	CR610888 full-leng
8	3564.5	98.1	2107	3	AK019460 Mus muscu
9	3549	97.7	2073	3	CR593724 full-leng

10	3549	97.7	2083	3	CR608666 full-leng
11	3530.5	97.2	2373	3	AK034755 Mus muscu
12	3455.5	95.1	2236	3	CR604639 full-leng
13	3321	91.4	2269	3	AK049743 Mus muscu
14	2568	70.7	2692	3	BC043904 Xenopus 1
15	1874	51.6	1379	3	AK014366 Mus muscu
16	1675.5	46.1	4729	3	HS0804802 Homo sapi
17	1672	46.0	1049	5	EX325713
18	1663	45.8	1045	5	BM909438
19	1656	45.6	1008	5	BM353313
20	1645.5	45.3	1058	4	BM802924
21	1629.5	44.9	1056	5	BQ059127
22	1629	44.8	1020	5	BM927718
23	1628	44.8	1015	5	BU841107
24	1628	44.8	1079	5	BM449774
25	1624.5	44.7	1062	5	BM909485
26	1624	44.7	962	5	BM443095
27	1622	44.6	1021	5	BM464988
28	1614	44.4	984	5	BM443496
29	1609.5	44.3	1040	5	BM399070
30	1608.5	44.3	1787	3	CR618632
31	1607	44.2	1027	5	BM428312
32	1607	44.2	1062	4	BM473356
33	1607	44.2	1068	5	BM382091
34	1604.5	44.2	1024	4	BM556024
35	1604	44.2	1024	5	BQ062138
36	1602	44.1	993	5	BM462635
37	1602	44.1	1087	5	BM398765
38	1601	44.1	1028	4	BM476095
39	1599	44.0	922	5	BM465100
40	1597	44.0	1013	5	BM367687
41	1589	43.7	1024	4	BM468711
42	1583	43.6	1031	4	BM477969
43	1574.5	43.3	1041	5	BM420692
44	1573	43.3	971	5	BM407826
45	1568	43.2	917	5	BM407419

## ALIGNMENTS

RESULT 1  
LOCUS CR604952 2157 bp mRNA linear HTC 21-JUL-2004  
DEFINITION full-length cDNA clone CS0D10107112 of Placenta Cot 25-normalized  
of Homo sapiens (human).  
ACCESSION CR604952.1 GI:50485759  
VERSION CR604952.1 GI:50485759  
KEYWORDS HTC; CDS; EST; CDNA.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens

REFERENCE  
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
TITLE 1 (bases 1 to 2157)  
JOURNAL Li, W.B., Gruber, C., Jessee, J. and Polyes, D.  
REMARK Unpublished  
Full-length cDNA libraries and normalization  
Contact: Feng Liang Email: fliang@lifetech.com URL:  
http://fulllength.invitrogen.com/Invitrogen Corporation 1600  
Faraday Avenue  
2 (bases 1 to 2157)

REFERENCE  
AUTHORS Genoscope.  
TITLE Direct Submision  
JOURNAL Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage :  
BP 191 91006 EVRY cedex - FRANCE (E-mail: seqref@genoscope.cns.fr  
- Web: www.genoscope.cns.fr)

COMMENT  
1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime  
end enriched, double-strand cDNA was digested with Not I and cloned  
into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library  
was normalized. Library was constructed by Life Technologies, a  
division of Invitrogen.

FEATURES  
source  
1..2157  
/organism="Homo sapiens"



Db	1961	AAATGATTAAGGCGACACCGTACGAGCGCAGATGCGCCCTAC	2008
Db	CR608375	2314 bp	mRNA
LOCUS	CR608375		linear
DEFINITION	full-length cDNA clone CSOD1039YJ24 of Placenta Cot 25-normalized of Homo sapiens (human).		HTC 21-JUL-2004
ACCESSION	CR608375		
VERSION	CR608375.1		
KEYWORDS	HTC; CNSLT_cDNA.		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
AUTHORS	1. (bases 1 to 2314)		
TITLE	L1.W.B., Gruber,C., Jeesee,J. and Polayes,D.		
JOURNAL	Full-length cDNA libraries and normalization		
REMARK	Unpublished		
CONTACT	Contact : Feng liang Email : fliang@life.technet.com URL :		
REFERENCE	http://fulllength.invitrogen.com/ Invitrogen Corporation 1600		
AUTHORS	Faraday Avenue		
TITLE	2. (bases 1 to 2314)		
JOURNAL	Genoscope.		
REMARK	Direct Submission		
COMMENT	Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : sequef@genoscope.cns.fr)		
FEATURES	1. .2314		
source	/organism="Homo sapiens"		
	/mol_type="mRNA"		
	/db_xref="taxon:9606"		
	/clone="CSOD1039YJ24"		
	/tissue_type="Placenta Cot 25-normalized"		
	/plasmid="pCMVSPORT_6"		
ALIGNMENT	Location/Qualifiers		
Alignment Scores:			
Pred. No.:	1.33e-212	Length:	2314
Score:	3633.00	Matches:	656
Percent Similarity:	100.00%	Conservative:	0
Best local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
Db:	3	Gaps:	0
US-10-791-017A-2	(1-656) x CR608375 (1-2314)		
Qy	1	MeAlaSeRThrAaPpYSeRThrTySeRgInAlaAlaGInGInGlyTySeRaLa	20
Db	41	ATGGCGTCCAGGATTACAGTACCTATGACCAAGTCCAGGCGACGACGACGCTACAGTCT	100
Qy	21	TyThAlaGInPProThGInGInGlyTyAlaGInThThGInAlATyGlyGInGInSer	40
Db	101	TACACCGCCGACCCCACTCAAGATATCACAGACCAACGCGCATATGCGCAACAAAGC	160
Qy	41	TyGlyThrTyGlyGInPProThRaPpYSeRThrThGInAlaGInThThAlaThr	60
Db	161	TATGGACCTATGACAGCCCACTATATCTCACTTACCCAGGCTCAAGACCACTGCAAC	220
Qy	61	TyGlyGInThAlATyAlAThSeSeTyTyGlyGInPProThGInTyTyThThPro	80
Db	221	TATGGGCGACGCGCTATGCAACTCTTATGACAGCGCTCCCACTGTTATATCACTCA	280
Qy	81	ThAlAPProGInAlATySeSeGInPProVaGInGInGlyTyTyGlyThThAlATyTaPpThr	100
Db	281	ACTGCGCCCGACGACATACGACGCTCTTCCAGGGGATGACATGAGTATGATATACC	340

QY	101	ThrThrAlaTrValTThrThThThGlnAlaSerValAlaIaGlnSerAlaValGlyThr	120
Db	341	ACCACTGCTAAGTACACCAACCAACCAAGGCTCTTATGCACTAGTCTGCATATGCACT	400
QY	121	GlnProAlaTrProAlaTrValGlyGlnGlnProAlaAlaThraAlaProThraTrProGln	140
Db	401	CAGCTGCTTATCCAGCTTATGGGCAAGACAGACAGCACTGCACCTCAACAAAGCCGAG	460
QY	141	AspGlyAsnLysProThrGlnLuthSerGlnProGlnSerSerThrGlnGlyTyrAsnGln	160
Db	461	GATGAAACAAAGCCCACTGAGACTAGTCAACCTCAACTGACAGGGGGTTTAAACACAG	520
QY	161	ProSerLeuGlyTyrGlyGlnSerAsnTyrSerTyrProGlnValProGlySerTyrPro	180
Db	521	CCAGCTTAGGATATGGACAGATCACTCAAGTATCCCAAGTACCTGGAGCTTACCC	580
QY	181	MetGlnProValThraAlaProProSerTyrProProThrSerTyrSerSerThrGlnPro	200
Db	581	ATGACAGCAGTCACTGCACCTCCATCCATACCCTCTACACAGCTATTCTCTACACAGCG	640
QY	201	ThiSerTyrAspGlnSerSerSerTyrSerGlnGlnAsnThrTyrGlyGlnProSerSerTyr	220
Db	641	ACTAGTTATGATCAGACAGATTACTCTCAGAGAAACCTATGGGCAACCAAGCAGCTAT	700
QY	221	GlyGlnGlnSerSerTyrGlyGlnGlnSerSerTyrGlyGlnGlnProProThrSerTyr	240
Db	701	GGACAGCAGATACCTATGTCACAAAGCAGCTATGGGCAAGCAGCTCCCACTAGTAC	760
QY	241	ProProGlnThrGlySerTyrSerSerGlnAlaProSerGlnTyrSerGlnGlnSerSer	260
Db	761	CCACCCCAATGTGATCTTACAGCCCAAGCTCAACTCAATTATACCAACAGACAGCAGC	820
QY	261	TyrGlyGlnGlnSerSerSerPheArgGlnAspHisProSerSerMetGlyValTyrGlyGln	280
Db	821	TACGGGCAAGCAGATTCAATCCGACAGACACCCCAAGTACAGGGGTATTATGGGCAAG	880
QY	281	GlnSerGlyGlyPheSerGlyProGlyGlnAsnArgSerMetSerGlyProAspAsnArg	300
Db	881	GAGTCTGAGGATTTTCCGACCAAGAGAGAACCGGAGCAGATGGCCCTGATTAACCGG	940
QY	301	GlyArgGlyArgGlyGlyPheAspArgGlyGlyMetSerArgGlyGlyArgGlyGly	320
Db	941	GGCAGGGGAAGAGGGGGATTTGATCTGTGAGGCATGACAGAGGTGGGGGGAGAGCA	1000
QY	321	ArgGlyGlyMetGlySerAlaGlyGlnArgGlyGlyPheAsnLysProGlyGlyProMet	340
Db	1001	CGGGTGGAAATGGGCAAGCGGTGAGACAGAGTGCTTCAATAAAGCCTGGTGGACCATG	1060
QY	341	AspGlnGlyProAspLeuAspLeuGlyProProValAspProAspGlnAspSerAspAsn	360
Db	1061	GATACAGAGCACAGACTCTTGATCTTGAGCCCACTGTAAATCCAGATGAAAGACTCTGACAAC	1120
QY	361	SerAlaLeuTyrValGlnGlyLeuAsnAspSerAlaThiLeuAspAspLeuAlaAspPhe	380
Db	1121	AGTCAATTATATGACAAAGATTAAAGACAGTGTACTAGATGATCTGGCAGACTTC	1180
QY	381	PheLeuGlnCysGlyValValLysMetAsnLysArgThrGlyGlnProMetLeuHisIle	400
Db	1181	TTTAAAGCAAGTGGGGTCTTTAAAGATGAACAAGAACTGGGCAACCAATCATCAATC	1240
QY	401	TyrLeuAspLysGlnThrGlyLysProLysGlyAspAlaThraValSerTyrGlnAspPro	420
Db	1241	TACCTGGAACAAGAAACAGAAAGCCCAAAGGCATGCCACAGTGTCTTATGAAGACCA	1300
QY	421	ProThrAlaLysAlaAlaValGlnTrpPheAspGlyLysAspPheGlnGlySerLysLeu	440
Db	1301	CCCACTGCAAAGCTGCGGGGAATGGTTGATGGGAAAGTTTCAAGGAGCAAACTT	1360
QY	441	LysValSerLeuAlaArgLysLysProProMetAsnSerMetArgGlyGlyLeuProPro	460
Db	1361	AAATCTCTCCCTGCTCGGAAGAACCTCTCAATGAACAGTATGGGGGTGTGTCTCCACCC	1420
QY	461	ArgGlnGlyArgGlyMetProProProLeuAspGlyGlyProGlyGlyProGlyPro	480

Db	1421	CGTAGGGCAGAGGCATGCCACCACTCCGTGGAGGTTCCAGGAGGCCACGAGAGTCTCT	1480
Qy	481	GLYGLYProMetGlyArgMetGlyGlyYArgGlyGlyYAspArgGlyGlyYPhenProProArg	500
Db	1481	GGGGGACCCAGAGGGGTCCGATGGGAGGCCGTGGAGAGATAGAGAGGGCTTCCCTCCAGA	1540
Qy	501	GLYProArgGlyYSerArgGlyYAsnProSerGlyGlyGlyYAsnValGlnHisArgAlaGly	520
Db	1541	GGACCCCGGGGTTCCCGAGGAAACCCCTCTGGAGAGGAACGTCACGACCGAGCTTGA	1600
Qy	521	AspTrpGlnCysProAsnProGlyCYsGlyYAsnGlnAsnPheAlaTrpArgThrGlnCys	540
Db	1601	GACCTGGCAGTGTCCCAATCCGGGTTGTGGAAACAGAACTTCGCTGGTAGAACAAGATGC	1660
Qy	541	AsnGlnCysIleuAspProIleProGlnGlyPheLeuProProPheProProProGly	560
Db	1661	AACCAAGTAAAGGCCCCCAAGCCCTGAAGGCTTCTCCGCAACCTTTCGCCCCCGGGT	1720
Qy	561	GlyYAspArgGlyYArgGlyGlyYProGlyYAsnMetArgGlyGlyYArgGlyGlyYLeuMetAsp	580
Db	1721	GGTATCGTGGCAGAGGTGGCCCTGGTGGTGCATGGCGGGAGGAAGAGTGGCTCATGAT	1780
Qy	581	ArgGlyGlyYProGlyGlyYMetPheArgGlyGlyYArgGlyGlyYAspArgGlyGlyYPheArg	600
Db	1781	CGTGTGTGTCGGGTGGATGATTTCAAGGTGGCCGTGTGGACAGAGGTGGCTTCGT	1840
Qy	601	GlyGlyYArgGlyYMetAspArgGlyGlyYPhenGlyGlyGlyYArgArgGlyGlyYProGlyGly	620
Db	1841	GGTGGCCGGGGCATTGACCGAGAGTGGCTTTGTGTGAAGAAACAGAGTGGCCCTGGGGGG	1900
Qy	621	ProProGlyYProLeuMetGlnGlnMetGlyGlyYArgArgGlyGlyYArgGlyGlyYProGly	640
Db	1901	CCCCCTGGACCTTTGATGATGAACAGATGGAGAAAGAAAGAGAGACGTGGAGGACCTGA	1960
Qy	641	LyseMetAspIleuArgGlyGlnHisArgGlnGlnArgArgAspArgProIleY	656
Db	1961	AAATATGATTAAGGAGACACCGTCGTAGAGCGCAGAGATGCGCCCTTAC	2008

RESULT 3	
CR625247	
LOCUS	
DEFINITION	CR625247 full-length cDNA clone CSDBE0057f1 of Placenta of Homo sapiens (human) .
ACCESSION	CR625247
VERSION	CR625247.1 GI:50506054
KEYWORDS	HTC; CNSLT cDNA.
SOURCE	Homo sapiens (human)
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS	Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
TITLE	Full-length cDNA libraries and normalization
JOURNAL	Unpublished
REMARK	Contact : Feng Liang Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com/ Invitrogen Corporation 1600 Faraday Avenue
REFERENCE	Genoscope.
AUTHORS	2 (bases 1 to 2159)
TITLE	Direct Submission
JOURNAL	Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
COMMENT	- Web : www.genoscope.cns.fr) 1st strand cDNA was primed with a NotI-oligo(dAT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies, a division of Invitrogen. location/Qualifiers 1..2159 /organism="Homo sapiens" /mol_type="mRNA"
FEATURES	
source	

```

/db_xref="taxon:9606"
/clone="CS0DE005YF11"
/tissue_type="Placenta"
/plasmid="pCMVSPORT_6"

```

Alignment Scores:	
Pred. No.:	9.5e-212
Score:	3618.50
Percent Similarity:	99.88%
Best Local Similarity:	99.88%
Query Match:	99.60%
DB:	3
Length:	215
Matches:	655
Conservative:	0
Mismatches:	0
Indels:	1
Gaps:	1

US-10-791-017A-2 (1-656) x CR625247 (1-2159)

QY	1	MeAlaSerThrAspIlyrSerThrTyrSerGlnAlaAlaAlaGlnGlnGlnIlyrSerAla	20
Db	46	ATGGGCTCCAGATTACAGTACTATAGCCAAAGCTGACGCCAGCGGCTTCACTGCT	105
QY	21	TyThrAlaGlnProthrgInglYTyraAlaGlnthrrhGlnAlaTyrgIyGlnInser	40
Db	106	TACACCGCCAGCCCACTCAAGAGATATATGACAGACCACCGAGCATATGGCAACAAAGC	165
QY	41	TyrgIyThrrTyrgIyGlnProthrrAspValSerTyrrThrrGlnAlaGlnthrrThraAlaThr	60
Db	166	TATGGAACCTATGAGCAGGCCCATGATGTCAAGCTATACCCAGGCTCAGACCACTGCAACC	225
QY	61	TyrgIyGlnthrrAlaTyraAlaThrSerTyrgIyGlnProProthrgIyTyrrThrrThrPro	80
Db	226	TATGGGAGAGACCGGCTATGCAACTTCTTATGGACAGCGCTCCACATGGTTATACTATCA	285
QY	81	ThrAlaProGlnAlaTyrsSerGlnProValGlnGlnIlyTyrgIyThGlnAlaTyraAspThr	100
Db	286	ACTGCCCCCGGAGCATACAGCAGCGCTCTCCAGGGGTATGGACATGGTCTTATATGATACC	345
QY	101	ThrThraAlaThrValThrThrThGlnAlaSerTyraAlaAlaGlnSerAlaTyrgIyThr	120
Db	346	ACCACGTGTAAGTACACCAACCCAGGCTCTCTATGCAAGCTCACTTGCATATGGCACT	405
QY	121	GlnProAlaTyrrProAlaTyrgIyGlnGlnProAlaAlaThrAlaProThrrArgProGln	140
Db	406	CAGCTGTCTTATCCAGCTTATGGGAGAGAGCAGCAGCAGCACTGACATTCACAGACCGCAG	465
QY	141	AspGlyAsnIlysProthrgInuThrSerGlnProGlnInserSerThrgIyGlnTyrrArgGln	160
Db	466	GATGGAACCAAGCCCATCTGAACATGATAACTCATCTAGCAGAGGGGTATCAACACAG	525
QY	161	ProSerLeuGlyTyrgIyGlnSerAsnTyrsSerTyrrProGlnValProGlySerTyrrPro	180
Db	526	CCGAGCGTAGATATGACAGAGATACATCAAGTTATCCCGAGGTACTGTGGAGCTACCCC	585
QY	181	MetGlnProValThrAlaProProSerTyrrProProthrrSerTyrrSerSerThrgInPro	200
Db	586	ATGCAAGCAGATCAGTCACTTCATCTCAACCTCTTACACAGTATTCCTTACACAGCCG	645
QY	201	ThrsertyraAspGlnSerSerTyrsSerGlnGlnAsnThrrTyrgIyGlnProSerSerTyrr	220
Db	646	ACTAGTTATGATCCAGACAGATTACTCTCAGCAGAACTATATGGGCAACCGAGCACTAT	705
QY	221	GlyGlnGlnInserSerTyrgIyGlnGlnInserSerTyrgIyGlnGlnProProthrrSerTyrr	240
Db	706	GGAAGAGCAGATACTATGTGTCAACAAAGCAGCTATGGGAGCAGACCTCTCCACTAGTTAC	765
QY	241	ProProGlnthrrGlyserTyrsSerGlnAlaProSerGlnTyrsSerGlnGlnInserSer	260
Db	766	CCACCCCAACTGATCTCTACAGCCCAAGCTCCAACTCAATATATAGCCAAAGCAGCAGAC	825
QY	261	TyrgIyGlnGlnInserSerPheArgGlnAspHisProSerSerMetGlyValTyrgIyGln	280
Db	826	TACGGGACGACAGATTCATTCGACAGGACCAACCCCACTAGCATGGGTGTTTATATGGCAG	885
QY	281	GlnSerGlyGlyPheSerGlyProGlyGlnAsnArgSerMetSerGlyProAspAsnArg	300



Db	886	GAGTCGGAGGATTTTCCGGACCAAGGAGAAACCGGACATGAGTGGCCCTGATTAACCGG	945
Qy	301	GIYARSGIYARSGIYGLYIYPHEAPARGIYGLYMETSERARGIYGLYARGIYGLYGLY	320
Db	946	GGCAGGGGAMAGGGGGGATTTGATCGTGAAGGACAGACAGAGTGCGGGGAGAGAGGA	1005
Qy	321	ARGGIYGLYMETGISYSEALAGIYGIUNARGIYGLYIYPHEANLYSPROGIYGLYIYPROMET	340
Db	1006	CGCGGTGGAAATGGGC--GCTGGAGAGGAGGTGGCTTCATTAATACCTGCTGGACCCCATG	1062
Qy	341	ASPGIUGIYPROAPBLEUASPLeUGIYPROPROVALAPPROAPSPGIUNAPSERASPAN	360
Db	1063	GATBAGGACCAAGATCTTGATCTAGGCCCACTGTAGATCCAGATGAAGACTCTGACAAAC	1122
Qy	361	SERALILEIYRVALGINGIYLEUANAAPSERVALIIRLEUASPAPLEUALAAPHHE	380
Db	1123	AGTCAATTTATGTACAGAGATTAAATGACAGTGTGACTTAAGATGATCTGGACAGACTTC	1182
Qy	381	PHELYSGINCYRGIVALVALLYIWMETANLYSAGTHIRGIYGLINPROMETIIHEIILE	400
Db	1183	TTTAAAGCATGTGGGGGTGTTAAGATGAAACAAGAACTGGGCAACCATGATTCACATCT	1242
Qy	401	TYRLEUAPLYBSGLUTHGIYLYSPROLYSGIYASPAIATRVASERTYRGUNAPPRO	420
Db	1243	THCCTGGACAAAGAAACAGAAAGCCCAAGGCGATGCCACAGTGTCTTATGAAAGACCA	1302
Qy	421	PROTHRALIYBALALALVALGILUTRPHENAPGIYLYSAPPHENGLIYSEIRYBLEU	440
Db	1303	CCCATGCGCAAGGCTGCCGTGGAAATGTTTGAATGGAAAGATTTTCAAGGGAGCAAACTT	1362
Qy	441	LYSVALSERLEUALAARGLYLYBPROBROMETANSEMETARGIYGLYLEUPROPRO	460
Db	1363	AAAGTCTCTCCCTTGCTCGGAMAGCCTCAATGAAACAGTATGCGGGGTGCTGCCACCC	1422
Qy	461	ARGIUGIYARSGIYMETPROBROBLEUARGIYGLYIYPROGIYGLYIYPROGIYGLYIYPRO	480
Db	1423	CTTAAGGGCAAGGCAATCCACACACTCGGTGAGGTCCAGAGAGGCCCAAGAGSTCTT	1482
Qy	481	GIYGLYIYPROMETGIYARGMETGIYGLYIYARGIYGLYIYASPARGIYGLYIYPHEPROPROARG	500
Db	1483	GGGGAGCCCAATGGGTGCGATGGAGAGCCGTGGAGAGATAGAGAGGCTTCCCTCCACGA	1542
Qy	501	GLYIYPROARGIYSEARARGIYASNPROSERGIYGLIYAEUNVALGINHIAARGHLAGIY	520
Db	1543	GGACCCCGGGGTTCCCGAGGGAAACCCCTCTGGAGGAGGAACGTGCACACACGAGCTGGA	1602
Qy	521	ASPTRPGINCYBPROAENPROGIYCYSELIVANGIUNANPHEALATTPARGTHGIUCYB	540
Db	1603	GACTGGCAGTGTCCCAATCCGGGTTGTGAAACCAAGAACTTCGCTGGAGAAACAGAGTGC	1662
Qy	541	ASNGINCYBLYBALAPPROLYSPROGIUGIYPHELEUPROPROPOBHEPROPROBROGIY	560
Db	1663	AACCAAGTITTAAGGCCCCCAAGCTCTAGAGGCTTCTCCCGCAACCTTTCCGCCCCCGGCT	1722
Qy	561	GLIYASPARARGIYARSGIYGLIYPROGIYGLIYMETARGIYGLIYARGIYGLIYLEUWETASP	580
Db	1723	GGTGATCGTGCAAGAGGTGGCCCTGTGTGCATGCGGGAGGAAGAGTGGCTCATGAT	1782
Qy	581	ARGGIYGLIYPROGIYGLIYMETPHEARGIYGLIYARGIYGLIYASPARGIYGLIYIYPHEARG	600
Db	1783	CGTGGTGTCCCGGGGAATGTCCAGAGTGTCCCTGTGTGAGACAGAGTGTGCTTCGCT	1842
Qy	601	GLIYGLIYARSGIYMETASPARGIYGLIYPHEGIYGLIYIYASPARSGIYGLIYIYPROGIYGLIY	620
Db	1843	GGTGGCCGGGCAATGACCCAGAGTGGCTTTGTGTGAGGAAGACAGAGTGGCCCTGGGGGGG	1902
Qy	621	PROBROGIYPROLEUMETGIUNMETGIYGLIYARGARGIYGLIYARGIYGLIYIYPROGIY	640
Db	1903	CCCCCTGGACCTTTGATGAGAACAGATGGAGGAAGAAAGAGGAGACGTGAGAGACCTGGA	1962
Qy	641	LYIWMETAPLYBSGLIYLIHIAARGIUNATRGATGAPATGPROTYR	656

DB	1963	AAATGATTAAGGAGACACCGTCAGAGGACAGATCGCCCTAC	2010
RESULT 4			
LOCUS	CR619493	2103 bp	mRNA linear HTC 21-JUL-2004
DEFINITION	full-length cDNA clone CSOD1076Y021 of Placenta Cot 25-normalized		
ACCESSION	CR619493.1	GI:50500300	
VERSION	HTC; CNSLT_CDNA.		
KEYWORDS	Homo sapiens (human)		
SOURCE	Homo sapiens		
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
REFERENCE	1 (bases 1 to 2103)		
AUTHORS	Li, W.B., Gruber, C., Jeejee, J. and Polayes, D.		
TITLE	Full-length cDNA libraries and normalization		
JOURNAL	Unpublished		
REMARK	Contact: Feng Liang Email: fliang@fitech.com URL: http://fulllength.invitrogen.com/ Invitrogen Corporation 1600 Faraday Avenue 2 (bases 1 to 2103)		
REFERENCE	Genoscope.		
AUTHORS	Direct Submission		
TITLE	Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage : BP 191 91106 EVRY cedex - FRANCE (E-mail: seque@genoscope.cns.fr)		
JOURNAL	Web : www.genoscope.cns.fr		
COMMENT	1st strand cDNA was primed with a NotI-oligo(dN) primer. Five primers end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies, a division of Invitrogen.		
FEATURES	Location/Qualifiers		
source	1..2103		
	/organism="Homo sapiens"		
	/mol_type="mRNA"		
	/db_xref="taxon:9606"		
	/clone="CSOD1076Y021"		
	/tissue type="Placenta Cot 25-normalized"		
	/plasmid="pCMVSPORT_6"		
ORIGIN			
Alignment Scores:			
Pred. No.:	1,51e-211	Length:	2103
Score:	3615.00	Matches:	652
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	99.50%	Indels:	0
DB:	3	Gaps:	0
US-10-791-017A-2 (1-656) x CR619493 (1-2103)			
QY	5	AspGlySerThrTyrSerGlnAlaAlaGlnGlnGlyTyrSerAlaTyrThrAlaGln	24
Db	1	GATTACAGTACTCTTATGACCAAGCTGACAGCGAGGAGCTTACAGTCTTACACCGCCAG	60
QY	25	ProThrGlnGlyTyrAlaGlnThrThrGlnAlaTyrGlnGlnGlnSerTyrGlyThrTyr	44
Db	61	CCCACTAGAGATATGACAGACCAACCGGCAATATGGGCAACAAGCTAATGAACTAT	120
QY	45	GlyAlnProThrAspValSerTyrThrGlnAlaGlnThrThrAlaThrTyrGlyGlnThr	64
Db	121	GGAAGGCCCACTGATGTCAGCTATATACCAAGGCTCAGACCACTGCACCTATGGGAGACC	180
QY	65	AlaTyrAlaThrSerTyrGlyGlnProProThrGlyTyrThrThrProThrAlaProGln	84
Db	181	GCCATATCAACCTCTTATGACACAGCTCCCACTGATATACTACTCCAACTGCCCCAG	240
QY	85	AlaTyrSerGlnProValGlnGlyTyrGlyThrGlyAlaTyrAspThrThrThrAlaThr	104
Db	241	GCAATACAGCCAGCTGTCACAGGAGATATGACCTGCTGCTATGATACCAACACCTGCTCA	300
QY	105	ValThrThrThrGlnAlaSerTyrAlaAlaGlnSerAlaTyrGlyThrGlnProAlaTyr	124

Db 301 GTACCAACCAACCGAGGCTCTTATGAGCTCAGTCTGATATGGCACTGACCTGCTTAT 360  
 Qy 125 ProAlaTyrGlyGlnGlnProAlaAlaThrAlaProThrArgProGlnAapGlyAenLys 144  
 Db 361 CCAAGCCCTATGGAGGAGCGAGCGAGCGACCTGACCTTCAAGACCGAGAGTGAAGAACAG 420  
 Qy 145 ProThrGlnThrSerGlnProGlnSerSerThrGlyYrAenGlnProSerLeuGly 164  
 Db 421 CCCACTAGACTACTCACTCACTCACTAGACAGAGGGGTTTACACAGAGCCCAAGCTTAGA 480  
 Qy 165 TyrGlyGlnSerAenTyrSerTyrProGlnValProGlySerTyrProMetGlnProVal 184  
 Db 481 TATGACAGAGTAACTACAGTTATCCCAAGTACTGGAGCTACCCCATGACAGCCAGTGC 540  
 Qy 185 ThrAlaProProSerTyrProProThrSerTyrSerSerThrGlnProThrSerTyrAap 204  
 Db 541 ACTGCACTCATCTCTCACTCTCTTACAGCTATCTCTTACACAGCCGAGCTAGTTATGAT 600  
 Qy 205 GlnSerSerTyrSerGlnGlnAenThrTyrGlyGlnProSerSerTyrGlyGlnGlnSer 224  
 Db 601 CAGAGCACTTACTTCAAGAGAACACCTATGGGCAACCGAGACTATGACAGCAGAT 660  
 Qy 225 SerTyrGlyGlnGlnSerSerTyrGlyGlnGlnProThrSerTyrProProGlnThr 244  
 Db 661 AGCTATGCTCAACAAAGCAGCTATGGGAGAGCTCCCACTAGTTACCAACCCCAACT 720  
 Qy 245 GlySerTyrSerGlnAapProSerGlnTyrSerGlnGlnSerSerTyrGlyGlnGln 264  
 Db 721 GGATCTTACAGCCAAAGCTCCAAAGTCAATATAGCCAAAGAGAGAGCTACGGGAGAG 780  
 Qy 265 SerSerPheArgGlnAapPhePheSerSerMetGlyValTyrGlyGlnGlnSerGly 284  
 Db 781 AGTTTCACTTCGACAGAGACACCCAGATGCAATGGGTTTATGGGCAAGAGTCTGAGGA 840  
 Qy 285 PheSerGlyProGlyGlnAapSerMetSerGlyProAapAenArgGlyAargGlyAarg 304  
 Db 841 TTTTCCGAGCAAGAGAGAACCGGAGCATGATGGCCCTGATTAACCGGAGGAGAGAA 900  
 Qy 305 GlyGlyPheAapArgGlyGlyMetSerSerGlyGlyAargGlyGlyAargGlyMet 324  
 Db 901 GGGGGAATTTATCTGAGAGCATGAGAGAGTGGGGAGAGAGAGAGAGAGAGAGAGATG 960  
 Qy 325 GlySerAaGlyGlnAargGlyGlyPheAapLysProGlyGlyProMetAapGlnGlyPro 344  
 Db 961 GCGAGCCCTGAGAGAGCAGGCTTCAATAGCTGCTGAGACCCATGATGAAGAGCCA 1020  
 Qy 345 AapLeuAapLeuGlyProProValAapProAapGlnAapSerAapAenSerAlaIleTyr 364  
 Db 1021 GATCTTATCTAGGCCCACTGTAGATCCAGATGAGACCTGACCAACAGTGCATTTAT 1080  
 Qy 365 ValGlnGlyLeuAapAapSerValThrLeuAapAapLeuAapPhePheLysGlnCys 384  
 Db 1081 GTACAAAGATTAATAATGACAGGTGACTCTAGATGATGTGGAGACTTCTTAAAGCACTGT 1140  
 Qy 385 GlyValValLysMetAenLysAargThrGlyGlnProMetIleHisIleTyrLeuAapLys 404  
 Db 1141 GGGGTTTAAAGTGAACAGAGAGAACTGGGCAACCCATGATCCACATCTCTGAGACAG 1200  
 Qy 405 GlnThrGlyLysProLysGlyAapAlaIleThrValSerTyrGlnAapProProThrAlaLys 424  
 Db 1201 GAAACAGAGAAAGCCCAAGGCGATGACAGAGTGCCTATGAAGACCCACCACTGCCAAG 1260  
 Qy 425 AlaAlaValAlaGlnThrPheAapGlyLysAapPheGlnGlnLysLeuLysValSerLeu 444  
 Db 1261 GCTGCGCTGAAATGTTTGAATGGGAAAGATTTTCAAGGAGCAAACTTAAAGCTCCCTT 1320  
 Qy 445 AlaArgLysLysProProMetAenSerMetAargGlyGlyLeuProProAargGlnGlyAarg 464  
 Db 1321 GCTCGAAGAAAGCTCCATGAACAGATATCGGGGTGTCTGCCACCCCTGAGGGAGCA 1380  
 Qy 465 GlyMetProProProLeuAargGlyGlyProGlyGlyProGlyGlyProGlyGlyProMet 484

Db 1381 GGCATCCACCAACCACTCCGTGAGAGTCCAGAGAGCCCAAGAGTCTCGGAGCAACCATG 1440  
 Qy 485 GlyArgMetGlyGlyLysArgGlyGlyLysAapArgGlyGlyPheProProAapGlyProAargLys 504  
 Db 1441 GGTGCAATGAGAGCCGTGAGAGATGAGAGAGGCTTCCCTCAAGAGAGAGAGAGAGAGT 1500  
 Qy 505 SerArgGlyAapProSerGlyGlyLysAenValGlnHisAargLysAapTyrGlnCys 524  
 Db 1501 TCCGAG 1560  
 Qy 525 ProAapProGlyCysGlyAenGlnAenPheAlaTyrArgThrGlyCysAenGlnCysGly 544  
 Db 1561 CCCAATCCGGGTGTGGAACCAAGACTTCGCTGAGAGAACAGAGTGCACACAGTGTAAAG 1620  
 Qy 545 AlaProLysProGlnGlyPheLeuProProProPheProProProGlyGlyAapAargLys 564  
 Db 1621 GCCCAAGCTGAAAGCTTCTCCGCAACCTTTCGCGCCCGGGGTGTATCGTGGC 1680  
 Qy 565 ArgGlyGlyProGlyGlyMetAargGlyGlyAargGlyGlyLeuMetAapArgGlyGlyPro 584  
 Db 1681 AGAGTGGCCCTGTGTGATGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1740  
 Qy 585 GlyGlyMetPheArgGlyGlyLysArgGlyGlyLysAapArgGlyGlyPheArgGlyGly 604  
 Db 1741 GGTGGAATGTTCAAGAGTGGCGGTGTGAGACAGAGAGTGGCTTCCGTGTGTGCGCGGGC 1800  
 Qy 605 MetAapArgGlyGlyPheGlyGlyGlyAargAargGlyGlyProGlyGlyProProGlyPro 624  
 Db 1801 ATGAGCCAGAGTGGCTTGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1860  
 Qy 625 LeuMetGlnGlnMetGlyGlyAargArgGlyGlyAargGlyGlyProGlyLysMetAapLys 644  
 Db 1861 TTGATGAAACAGATGAG 1920  
 Qy 645 GlyGlnHisArgGlnGlnAargAargAapArgProTyr 666  
 Db 1921 GCGAGACACCTGTCAAGAGCGCAGAGATCGGCCCTAC 1986  
 RESULT 5  
 CR602561 2070 bp mRNA linear HTC 21-JUL-2004  
 LOCUS Full-length cDNA clone CS0D1068YN3 of Placenta Cot 25-normalized  
 DEFINITION of Homo sapiens (human).  
 ACCESSION CR602561 GI:50483368  
 VERSION CR602561.1  
 KEYWORDS HTC; cDNA; cDNA.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 REFERENCES Bukaryoca; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 AUTHORS Li, W.B., Gruber, C., Jesse, J., and Polayes, D.  
 TITLES Full-length cDNA libraries and normalization  
 JOURNAL Unpublished  
 REMARK Contact: Feng Liang Email: fliang@life.techn.com URL:  
 http://fulllength.invitrogen.com/Invitrogen Corporation 1600  
 Faraday Avenue  
 2 (bases 1 to 2070)  
 REFERENCE Genoscope.  
 AUTHORS Direct Submissions  
 TITLES Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage :  
 JOURNAL BP 191 91006 EVRY cedex - FRANCE (E-mail: seqref@genoscope.cns.fr  
 COMMENT 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime  
 end enriched, double-strand cDNA was digested with Not I and cloned  
 into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library  
 was normalized. Library was constructed by Life Technologies, a  
 division of Invitrogen.  
 FEATURES Location/Qualifiers  
 1..2070  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"

/clone="CS0DI068YN23"  
/cissue="Placenta Cot 25-normalized"  
/plasmid="pCMVSPORT\_6"

## ALIGNMENT SCORES:

Pred. No.:	1,15e-210	Length:	2070
Score:	3600.50	Matches:	651
Percent Similarity:	99.85%	Conservative:	0
Best Local Similarity:	99.85%	Mismatches:	0
Query Match:	99.11%	Indels:	1
		Gaps:	1

US-10-791-017a-2 (1-656) x CR602561 (1-2070)

QY 5 AspTyrSerThrTyrSerGlnAlaAlaGlnGlnGlnTyrSerAlaTyrThrAlaGln 24  
Db 1 GATTACAGTACCTTATGACCAAGCTGACGCGCAGGCGCTTACAGTGTTCACCGCCAG 60  
QY 25 ProThrGlnGlnTyrAlaGlnThrThrGlnAlaTyrGlnGlnGlnSerTyrGlnTyr 44  
Db 61 CCACCTCAAGAGATGACAGACCAACGCGCATATGGGCAACCACTATGGAACTAT 120  
QY 45 GlnGlnProThrAspValSerTyrThrGlnAlaGlnThrThrAlaThrTyrGlnThr 64  
Db 121 GGACAGCCCACTGATGTCACTATACCAAGCTCAAGACCACTGACACTATGGCAGACC 180  
QY 65 AlaTyrAlaThrSerTyrGlnGlnProProThrGlnTyrThrThrProThrAlaProGln 84  
Db 181 GCGTATGCACTTATATGACAGCTCCCACTGTTATATCTACCACTGCGCCCCAG 240  
QY 85 AlaTyrSerGlnProValGlnGlnTyrGlnTyrGlnAlaTyrAspThrThrThrAlaThr 104  
Db 241 GCATACAGCAGCAGCTGTCCAGGGGTATGCGACTGCTTATATGATACCACTGCTACA 300  
QY 105 ValThrThrThrGlnAlaSerTyrAlaAlaGlnSerAlaTyrGlnThrGlnProAlaTyr 124  
Db 301 GTACACCAACCCAGGCTCTCTATGACCTCAAGCTCATGATATGCACTCAAGCTCTAT 360  
QY 125 ProAlaTyrGlnGlnGlnProAlaAlaThrAlaProThrArgProGlnAspGlyAsnLys 144  
Db 361 CCAAGCTATGGGAGCAGCAGCAGCAGCAGCTCAAGACCGGAGATGAGAAACAG 420  
QY 145 ProThrGlnThrSerGlnProGlnSerSerThrGlnTyrGlnGlnProSerLeuGly 164  
Db 421 CCACAGAGACTGTCAACTCACTCATATGACAGGGGTATCAACACCGCAGCTGAGA 480  
QY 165 TyrGlnGlnSerAlaTyrSerTyrProGlnValProGlySerTyrProMetGlnProVal 184  
Db 481 TATGACAGAGTATCTACAGTTATCCCAAGGTACTGAGAGCTACCCATGACAGCTC 540  
QY 185 ThrAlaProProSerTyrProProThrSerTyrSerSerThrGlnProThrSerTyrAsp 204  
Db 541 ACTGACCTCATCTCTACCTCCCTACAGGCTATTCCTACACAGCGGACTAGTTATGAT 600  
QY 205 GlnSerSerTyrSerGlnGlnAlaThrTyrGlnGlnProSerSerTyrGlnGlnSer 224  
Db 601 CAGAGCGTATCTCTCAGCAGAACCACTATGGGCAACCGAGCACTATGACAGCAGT 660  
QY 225 SerTyrGlnGlnGlnSerSerTyrGlnGlnProProThrSerTyrProProGlnThr 244  
Db 661 AGCTATGCTCAACCAAGCAGCTATGGGAGAGCTCCCACTAGTATCCCACTCAACT 720  
QY 245 GlnSerTyrSerGlnAlaProSerGlnTyrSerGlnGlnSerSerSerTyrGlnGln 264  
Db 721 GGATCCCAAGCAGCAAGCTCCAGTCAATATAGCAACAGAGCAGCACTACGGGAGCAG 780  
QY 265 SerSerPheArgGlnAspAlaProSerSerMetGlyValTyrGlnGlnGlnSerGly 284  
Db 781 AGTTCAATCCGACAGGACCAACCCAGATGATGGTGTATATGGGAGGATCTGGAGGA 840  
QY 285 PheSerGlnProGlnGlnAlaArgSerMetSerGlnProAspAlaValArgGlyArg 304

Db 841 TTTTCCGACGACGAGAGAAACCGAGCATAGTGCCTCTATTAACCGGGGACGAGGAGA 900  
QY 305 GlnGlyPheAspArgGlnGlyMetSerArgGlnGlyArgGlnGlyArgGlnGlyMet 324  
Db 901 GGGGGAATTTATGCTGAGGACATGACAGAGTGGGCGGAGGAGGACCGGTGATG 960  
QY 325 GlnSerAlaGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 344  
Db 961 GGC--GCTGAGAGCAGAGGTGCTTCAATAGCTGTGTGACCCATGATGAAGGACA 1017  
QY 345 AspLeuAspLeuGlnProProValAspProAspGlnAspSerAspAsnSerAlaIleTyr 364  
Db 1018 GATCTTATCTTAGGCCCACTGTATGATCCAGATGAAGACTCTGACCAAGTGCATTTAT 1077  
QY 365 ValGlnGlnLeuAsnAspSerValThrLeuAspAspLeuAlaAspPheLeuGlnCys 384  
Db 1078 GTACAAAGATTAATATGACATGACCTCTAGATGATCTGACAGCTCTTTAAGCACTGT 1137  
QY 385 GlnValValLeuMetLeuArgThrGlnGlnProMetIleHisIleTyrLeuAspLys 404  
Db 1138 GGGGTGTTAAGATGAACAAAGAACTGGGCAACCCATGATCCACATCTACTGGACAAG 1197  
QY 405 GlnThrGlnLysProLysGlnAspAlaThrValSerTyrGlnLysProProThrAlaLys 424  
Db 1198 GAACAGAGAAAGCCCAAGGCGATGCCACAGTCTCTATGAAGACCAACCACTGCCAAG 1257  
QY 425 AlaAlaValGlnTyrPheAspGlnLysAspPheGlnLysSerLysLeuValSerLeu 444  
Db 1258 GTCGCGTGAATATGTTGATGGAAAGATTTTCAAGGAGAGCAAACTTAAGTCTCCCTT 1317  
QY 445 AlaArgLysLysProProMetLeuSerMetArgGlnGlnLeuProProAlaGlnArg 464  
Db 1318 GCTCGAAGAAAGCTCCAAATGAACAGTATGCGGGGTGTGTCACCCCTGAGGGAGA 1377  
QY 465 GlnMetProProProLeuAspArgGlnGlnProGlnGlnLysProGlnGlnProMet 484  
Db 1378 GGCATGCCACCACTCCTGTGAGGTCCAGAGGCCCAAGAGTCTGTGGGAGCCATG 1437  
QY 485 GlnArgMetGlnLysArgGlnGlnLysAspArgGlnLysPheProProAlaArgLys 504  
Db 1438 GTCGCGATGGAGGCGCTGAGAGATAGAGAGGCTTCCCTCAAGAGACCCCGGGGT 1497  
QY 505 SerArgLysAspProSerGlnGlnLysAsnValGlnHisArgAlaGlnAspTyrGlnCys 524  
Db 1498 TCCCGAGGAAACCTCTGTGAGAGAAAGCTCAGACCGACCTGAGACTGCTGT 1557  
QY 525 ProAspProGlnCysGlnLysGlnAlaAspPheAlaTyrArgThrGlnCysAsnGlnCysLys 544  
Db 1558 CCGATCCGGGTGTGTGAACCAAGAACTTGCCTGAGAAACAGATGCCAAGCTGTAAAG 1617  
QY 545 AlaProLysProGlnGlnPheLeuProProProPheProProProGlnLysAspArgLys 564  
Db 1618 GCCCAAAAGCTGAAGGCTTCTCCGCCACCTTTCCGCCCCGGGTGTGATCGTGGC 1677  
QY 565 ArgGlnGlnProGlnGlnMetLysArgGlnGlnLysLeuMetAspArgGlnGlnPro 584  
Db 1678 AAGAGTGGCCCTGTGTGACGCGGAGAGAAAGTGGCTCTATGATCTGTGTGTGCC 1737  
QY 585 GlnGlnMetPheArgGlnGlnLysArgGlnLysAspArgGlnLysPheArgGlnGln 604  
Db 1738 GGTGGAATGTTCAAGAGTGGCTGTGTGAGAACAGAGTGTGCTTCCGTGTGGCCGGGGC 1797  
QY 605 MetAspArgGlnGlnPheGlnGlnGlnLysArgGlnGlnLysProGlnGlnProProGln 624  
Db 1798 ATGACCGAGGTGGCTTTGTGTGAGAAAGCAGAGTGGCCCTGCGGGGGCCCCCGAGCT 1857  
QY 625 LeuMetGlnGlnMetGlnGlnLysArgGlnGlnLysArgGlnGlnLysProGlnLysMetAspLys 644  
Db 1858 TTGATGAACAGATGGAGAGAAAGAGAGAGCGTGGAGGACCTGGAAATGAGATTA 1917  
QY 645 GlnGlnHisArgGlnGlnArgArgAspArgProTyr 656  
Db 1918 GCGGAGCACCTCAAGAGCGGAGAGATCGGCTTAC 1953

LOCUS	CR594868	2093 bp	mRNA	linear	HTC 21-JUL-2004				
DEFINITION	Full-length cDNA clone CS0D1064YE15 of Placenta Cot 25-normalized of Homo sapiens (human).								
ACCESSION	CR594868								
VERSION	CR594868.1	GI:50475675							
KEYWORDS	HTC; cNSLT cDNA.								
SOURCE	Homo sapiens (human)								
ORGANISM	Homo sapiens								
REFERENCE	Mumtazota, Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Eumalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.								
AUTHORS	Li, W.B., Gruber, C., Jeesee, J. and Polayes, D.								
TITLE	Full-length cDNA libraries and normalization								
JOURNAL	Unpublished								
REMARK	Contact : Feng Liang Email : fliang@lifetech.com URL : <a href="http://fulllength.invitrogen.com/">http://fulllength.invitrogen.com/</a> Invitrogen Corporation 1600 Faraday Avenue								
REFERENCE	2 (bases 1 to 2093)								
AUTHORS	Genoscope.								
TITLE	Direct Submission								
JOURNAL	Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage ; BP 191 91006 Evry cedex - FRANCE (E-mail : <a href="mailto:seqref@genoscope.cns.fr">seqref@genoscope.cns.fr</a> )								
COMMENT	Web : <a href="http://www.genoscope.cns.fr">www.genoscope.cns.fr</a> 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five primers end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies, a division of Invitrogen.								
FEATURES	Location/Qualifiers								
source	1..2093								
	/organism="Homo sapiens"								
	/mol_type="mRNA"								
	/db_xref="taxon:9606"								
	/clone="CS0D1064YE15"								
	/tissue_type="Placenta Cot 25-normalized"								
	/plasmid="pCMVSPORT_6"								
ORIGIN									
Alignment Scores:									
Prod. No.:	1.16e-210	Length:	2093						
Score:	3600.50	Matches:	651						
Percent Similarity:	99.85%	Conservative:	0						
Best Local Similarity:	99.85%	Mismatches:	0						
Query Match:	99.11%	Indels:	1						
DB:	3	Gaps:	1						
US-10-791-017A-2 (1-656) x CR594868 (1-2093)									
Oy	5	AspTyrSerTrrTyrSerGlnAlaAlaAlaGlnGlnGlnTyrSerAlaTrrThrAlaGln	24						
Db	1	GATTACGACTCTTATGACCAAGCTGCGACGCGACGAGGAGCTCAAGTCTTACCAACGCCCG	60						
Oy	25	ProThrgInglYTyraGlnThrThrgInlaaTyrGlyGlnGlnSerTyrGlyThrTyr	44						
Db	61	CCCACTCAAGGATATGCAACAACCCAGGAGCATATGGCAACAAGGTATGGAACGAT	120						
Oy	45	GlyGlnProThraaPValSerTyrThrgInaGlnThrAlaAlaThrTyrGlyGlnThr	64						
Db	121	GGACAGCCCACTGATGTCAGCTAATACCAAGGCTCAGACCTGCAACCTATGAGGACAGCC	180						
Oy	65	AlaTyrAlaThrSerTyrGlyGlnProProThrgInGlyTrrThrThrProThraAProGln	84						
Db	181	GCCATATGCACTTCTTATGACAGGCTCCCACTGTTATCTACTCCACTGCCCCCAG	240						
Oy	85	AlaTyrSerGlnProValGlnGlyTyrGlyThrgInaTyrAspThrThrAlaThr	104						
Db	241	GCATATACAGCCAGCGCTGCAGAGGGATAGGACCTGGTCTTATGATACACACAGCTGATCA	300						
Oy	105	ValThrThrThrgInaIaSerTyrAlaAlaGlnSerAlaTyrGlyThrgInProAlaTyr	124						

[illegible]





OY	305	GLYGLYPHeaPARG1YGLYMeSeRaRg1YGLYARSG1YGLYGLYARGLYGLYMeT	324
Db	901	GGGGATTTTATCGTGGAGCATGAGCAGAGGTGGCGGGAGAGAGCGCGGTGAATG	960
OY	325	GLYSeRa1aG1YGLUaRGLYGLYPHeaenLYPProG1YGLYPProMeAmpGLUg1YPro	344
Db	961	GGC---GCTGGAGAGCGAGGTGGCTTCAATAAGCCCTGGTGAACCCATGGATGAAGACCA	1011
OY	345	AaPLeuAaPLeuGLYProProV1a1aPProAaPGLUaAaPSeRaAaPSeRa1a1eTYr	364
Db	1018	GATCTTGATCTAAGGCCACCTGTMAATCCAGATGAAGACTTCGAACAACGTCAATTAT	1077
OY	365	Val1G1NG1LYLeuAaAaPSeRa1ThrLeuAaAaPLeuA1AaPPhaPheLYSG1nCYs	384
Db	1078	GTCACAGATTTAAATGACAGTGTGACTCTAGATGATCTGGCAGACTTCTTTAAGCAAGT	1133
OY	385	GLYVal1LYMeCaAaLYaRgThrg1YGLInProMeL1eH1a1eTYrLeuAaPYs	404
Db	1138	GGGGTGTTTAAGATGAACAAGAGAACTGGGCAACCATGATCAATCACTTGAACAAG	1199
OY	405	GLUThrg1LYsPProLYaAaP1aThrValSeTYrGLUaAaPProThra1aLYs	424
Db	1198	GAAACAGAGAAAGCCCAAGGCGCATGCCACAGTGTCTTAAGAAAGCCACCTCCAG	1257
OY	425	Ala1aVal1GLUTrpPheAaP1LYyAaAPheG1NG1SeTYsLeuLYaValSe1u	444
Db	1258	GCTGCGGTGAATGGTTTGATGGAAAGATTTTCAAGGAGCAACTTAAATCTTCCTT	1311
OY	445	AlaArgLYaLYPProPMeCaAaSeRaSeRaArg1YGLYLeuPProAaRGLUg1YArg	464
Db	1318	GCTCGGAAGAAAGCTTCCAATGAACGTATGCGGGGTGGTCTGCCACCCGTTAGGGCAGA	1377
OY	465	GLYMeTProPProPProLeuAaRGLYGLYProG1YGLYProG1YGLYProG1YGLYProMeT	484
Db	1378	GGCATGCAACCACTCCGTGAGAGTTCAGAGAGGCCAGAGAGTCTGGGGAGACCCATG	1433
OY	485	GLYARMeTGLYGLYARGLYGLYLAaPARGLYGLYPheProP1aRGLYProARGLY	504
Db	1438	GGTGGCATAGGGAGCGCGTGGAGAAATAGAGAGGCTTCCCTCAAGAGAGACCCGGGT	1497
OY	505	SeRaRg1YLAaPProSeRGLYGLYGLYLAaenVal1G1NH1aRGLaGLYLAaPTTGLCYs	524
Db	1498	TCCCCAGAGAAACCCCTCTGAGAGAGAAACCTCCAGCACCGAGCTGGAGACTGGCAGTGT	1557
OY	525	ProAaPProG1YCYsG1YLAaNG1LAaNPheA1aTPARgThGLUaCYaAaGL1nCYeLYs	544
Db	1558	CCCAATCCGGGTTTGTAAGAACCAAGACTTCCGCTGAAGAACAGGTGCAACCAAGTGAAG	1611
OY	545	AlaPProLYProGLUg1YPhaLeuPProPProPheProPProG1YGLYLAaPARGLY	564
Db	1618	GCCCCAAAGCCTGAAGGCTTCTCCCGCAACCTTTCGCGCCCGGGGTGTATGTGGC	1677
OY	565	ARGLYGLYProGLYGLYMeCaRGLYGLYARGLYGLYGLYLeuMeCaAaPARGLYGLYPro	584
Db	1678	AGAGGTGGCCCTGGTGGCATGGGGAGAGAAAGGTGGCTCATGATCGTGTGTCTCC	1733
OY	585	GLYGLYMeTPhaARGLYGLYARGLYGLYLAaPARSG1YGLYPheARGLYGLYARGLY	604
Db	1738	GGTGAATGTTCAAGAGTGGCCGTGTGTGAACAAGAGTGGCTTCCGTGTGGCCGGGGC	1797
OY	605	MeCaPARARGLYGLYPheGLYGLYGLYARARGLYGLYProGLYGLYProPProGLYPro	624
Db	1798	ATGAGCCAGAGGTGGCTTGTGTGTGAAGAAAGAGGTGGCCCTGGGGGGGCCCTTGAGCT	1857
OY	625	LeuMeTGLUg1MeCaLYGLYARARGLYGLYARGLYGLYProGLYLYyMeCaAaLYs	644
Db	1858	TTGATGGAACAGATGGAGAGAAAGAGAGACGTGGAGGACCTGGAATAATGATATAA	1917
OY	645	GLYGLUTh1aRSG1GLUaYARARGAaPARGProTYr	656
Db	1918	GGCGAGCACCTTCAGAGAGCGCAAGATGTGGCCCTTAC	1953

RESULT 8	
AKO19460	2107 bp mRNA linear HTC 03-APR-2000
LOCUS	
DEFINITION	Mus musculus 18 days pregnant adult female placenta and extra embryonic tissue cDNA, RIKEN full-length cDNA library, clone:3830417B1 product:Ewing sarcoma homolog, full insert sequence.
ACCESSION	AKO19460
VERSION	AKO19460.1 GI:12859676
KEYWORDS	HTC; CAP trapper.
SOURCE	Mus musculus (house mouse)
ORGANISM	Eumetazoa; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE	
AUTHORS	Carninci, P. and Hayashizaki, Y.
TITLE	High-efficiency full-length cDNA cloning
JOURNAL	Meth. Enzymol. 303, 19-44 (1999)
MEDLINE	99279253
PUBMED	10349636
REFERENCE	
AUTHORS	
TITLE	
JOURNAL	Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Komno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y. Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE	20499374
PUBMED	11042159
REFERENCE	
AUTHORS	
TITLE	
JOURNAL	Shibata, K., Itoh, M., Aizawa, K., Nagaka, S., Sasaki, N., Carninci, P., Komno, H., Akiyama, T., Nishi, K., Kitamura, T., Tashiro, H., Itoh, M., Suni, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kasaiwagi, K., Fujikake, Y., Inoue, K., Togawa, Y., Iizawa, M., Ohara, E., Watabiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kita, A. and Hayashizaki, Y. RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer Genome Res. 10 (11), 1757-1771 (2000)
MEDLINE	20530913
PUBMED	11076861
REFERENCE	
AUTHORS	
TITLE	
JOURNAL	The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium. Functional annotation of a full-length mouse cDNA collection Nature 409, 685-690 (2001)
MEDLINE	
PUBMED	
REFERENCE	
AUTHORS	
TITLE	
JOURNAL	The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team. Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs Nature 420, 563-573 (2002)
MEDLINE	
PUBMED	
REFERENCE	
AUTHORS	
TITLE	
JOURNAL	Adachi, J., Aizawa, K., Akahira, S., Akimura, T., Arai, A., Bono, H., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Fukunishi, Y., Furuno, M., Hanagaki, T., Hara, A., Hayatsu, N., Hiramoto, K., Hitachi, T., Horii, F., Imotani, K., Ishii, Y., Itoh, M., Izawa, M., Kanekawa, T., Kato, H., Kawai, J., Koike, Y., Komno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Nishi, K., Nomura, K., Numazaki, R., Ono, M., Okazaki, Y., Okido, T., Owa, C., Saito, H., Salto, R., Sakai, C., Sakai, K., Sano, H., Sasaki, D., Shibata, K., Shibata, Y., Shingawara, A., Shiraki, T., Sobabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F., Tanaka, T., Tejima, Y., Toya, T., Yamamura, T., Yasunishi, A., Yoshida, K., Yoshino, M., Mizumatsu, M. and Hayashizaki, Y.
MEDLINE	
PUBMED	
REFERENCE	
AUTHORS	
TITLE	
JOURNAL	Direct Submission Submitted (18-AUG-2000) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsukuba-shi, Yokohama Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.jp, URL:http://genome.gsc.riken.jp/, Tel:81-45-503-9222, Fax:81-45-503-9216)





```

Db 1274 CCAAGTCGCAAGGCTGCGTGGATGTTTGGATGGAAAGATTTTCAAGGAGAACTT 1333
Qy 441 LyvA1SerLeuA1aaaglylybProProMeCAsnSerMeArGlylylLeuProPro 460
Db 1334 AAAGTGTCTTGTGCCCCAAAGAAAGCCCTCCATGAAACGATGCGGGAGGCAATGCCACT 1393
Qy 461 ArgGluGlyArgGlyMetProProProLeuArgGlylylProGlylylProGlylylPro 480
Db 1394 CGTGAAGGCGAGGGGATGCCACCACTTCTGTGAGAGTCTGTGTGCCCCAGAGAGCCCT 1453
Qy 481 GlylylProMetGlyArgMetGlylylArgGlylylAAsparGlylylPheProProArg 500
Db 1454 GGAAGAACCAATGGGTGCGATGGAGGCGCTGGAGAGACAGAGGGGCGCTCCCTCCAAAG 1513
Qy 501 GlyProArgGlySerArgGlyAAsnProSerGlylylGlyAAsnValGlnHisArgAlaGly 520
Db 1514 GGGCCCCGAGGCTCCAGAGAAACCCCTCTGGAGAGAAATGTCAGCACCGAGCTGGA 1573
Qy 521 AspTArgGlnCysProAAsnProGlylylCysGlyAAsnGlnAAsnHea1ATPArgThrGlyCys 540
Db 1574 GACTGGCAGTGTCCCATCCGGGCTGTGAAACCAAGAACTTCCGTTGGAAGACAAATGC 1633
Qy 541 AAsnGlnCysAlaYsAlaProlybProGlylylPheLeuProProPheProProProGly 560
Db 1634 AACCAAGTAAAGGCCCCTAAGCCGAGAGGCTTCTCCGCAACCTTCCACCTCCGGGT 1693
Qy 561 GlyAAsparGlyArgGlylylProGlylylMetArgGlylylArgGlylylLeuMetAAsp 580
Db 1694 GGGATGTGTGACAGAGGTGGCCCTGTGTGAGATGGAGAGAAAGAGAGACCTCATGAGAC 1753
Qy 581 ArgGlylylProGlylylMetPheArgGlylylArgGlylylAAsparGlylylPheArg 600
Db 1754 CGTGTGTCTCTCGAGGAATCTTCAGAGTGTGCGAGAGTGGAGACAGAGAGCTTCCGA 1813
Qy 601 GlylylArgGlyMetAAsparGlylylPheGlylylGlylylArgGlylylProGlylyl 620
Db 1814 GGTGGCGGTGAATGACACGAGGTGCTTGTGTGAGAGAAAGACAGAGTGTCTCGGGGG 1873
Qy 621 ProProGlyProLeuMetGlnGlnMetGlylylArgGlylylArgGlylylProGlylyl 640
Db 1874 CTTCTGTGACCTTTAATGAGAACAGATGGAGAGAAAGAGCGGACCTGAGAGACCTGGG 1933
Qy 641 LyMetAAsparGlylylAAsnArgGlnGlnArgAAsparGlylylPro 656
Db 1934 AAATGATTAAGCGAGACACCTCGAGGAACGAGAGACCGGCTTAC 1981

RESULT 9
CR593724 2073 bp mRNA linear HTC 21-JUL-2004
LOCUS full-length cDNA CS0D1011Y005 of Placenta Cot 25-normalized
DEFINITION of Homo sapiens (human).
ACCESSION CR593724
VERSION CR593724.1 GI:50474531
KEYWORDS HTC; CNSLT cDNA.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1. (bases 1 to 2073)
AUTHORS Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished
REMARK Contact: Feng Liang Email: fliang@lifetech.com URL:
http://fulllength.invitrogen.com/InvitrogenCorporation1600
Faraday Avenue
2. (bases 1 to 2073)
REFERENCE Genoscope.
AUTHORS Direct Submission
TITLE Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage :
JOURNAL BP 191 91006 Evry cedex - FRANCE (E-mail: seqref@genoscope.cns.fr
COMMENT - Web : www.genoscope.cns.fr)
1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime

```

```

end enriched, double-strand cDNA was digested with Not I and cloned
into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
was normalized. Library was constructed by Life Technologies, a
division of Invitrogen.
FEATURES
source
Location/Qualifiers
1..2073
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0D1011Y005"
/tissue_type="Placenta Cot 25-normalized"
/plasmid="pCMVSPORT_6"
ORIGIN
Alignment Scores:
Pred. No.: 1,65e-207 Length: 2073
Score: 3549.00 Matches: 639
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 97.69% Indels: 0
DB: 3 Gaps: 0
US-10-791-017A-2 (1-656) x CR593724 (1-2073)
Qy 18 TySerAlaTyThrAlaGlnProThrGlnGlyTyThrAlaGlnThrThrAlaTyGly 37
Db 3 TACAGTCTTACACCCGCCAGCCCACTCAAGATATGACAGACCAACCCAGGATATGG 62
Qy 38 GlnGlnSerTyGlyThrTyGlyGlnProThrAAsparValSerTyThrGlnAlaGlnThr 57
Db 63 CAACAAAGCTATGAACTATGAGACGCCACATGATGATGATGATGATGATGATGATGATGAT 122
Qy 58 ThrAlaThrTyGlyGlnThrAlaTyThrAlaThrSerTyGlyGlnProThrGlyTyThr 77
Db 123 ACTGCAACCTATGGAGACCGCTATGACACTTCTTATGAGACGCTCCACCTGTTAT 182
Qy 78 ThrThrProThrAlaProGlnAlaTySerGlnProValGlnGlyTyThrGlyAla 97
Db 183 ACTACCTCAATGCCCCCAGGATACAGCCAGCTGTCCAGGGGATATGGCACTGGGTCT 242
Qy 98 TyAAspThrThrAlaThrValThrThrGlnAlaSerTyThrAlaGlnSerAla 117
Db 243 TATGATACCAACCACTGCTCAATGACACACACCAAGGCTCTTATGAGCTCTGCA 302
Qy 118 TyGlyThrGlnProAlaTyProAlaTyGlyGlnGlnProAlaThrAlaProThr 137
Db 303 TATGCACTGAGCTGCTTATCCAGCTATGGAGACAGCACAGCACAGCACAGCACAGCTTACA 362
Qy 138 ArgProGlnAAsparGlyAAsnLybProThrGlnThrSerGlnProGlnSerSerThrGlyly 157
Db 363 AGACCGCAGAGATGGAACAGGCCACTGAGACTGATCAACTCAACTCAACTCAACTCAACTCA 422
Qy 158 TyAAsnGlnProSerLeuGlyTyGlyGlnSerAAsnTySerTyProGlnValProGly 177
Db 423 TACAAACAGCCCGCTAGATGATGAGACAGAGATTAACAGTATATCCCGAGTACCTGGG 482
Qy 178 SerTyProMetGlnProValThrAlaProProSerTyProProThrSerTySerSer 197
Db 483 AGCTACCCATGAGCAGCAGTCACTGCACTCACTCACTCACTCACTCACTCACTCACTCT 542
Qy 198 ThrGlnProThrSerTyAAsnGlnSerSerTySerGlnGlnAAsnThrTyGlyGlnPro 217
Db 543 ACACAGCCGATGATGATATGATGACAGCACTTCTCAAGAGAAACCTATGGGCAACG 602
Qy 218 SerSerTyGlyGlnGlnSerSerTyGlyGlnGlnSerSerTyGlyGlnGlnProPro 237
Db 603 AGCAGCTATGACAGAGAGTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 662
Qy 238 ThrSerTyProProGlnThrGlySerTySerGlnAlaProSerGlnTySerGlnGln 257
Db 663 ACTAGTACCAACCCCAAGCTGATCTTACAGCAGCAAGCTCCAAATATATGACCAAG 722
Qy 258 SerSerSerTyGlyGlnGlnSerSerPheArgGlnAAsparSerSerMetGlyVal 277

```





**TITLE** RIKEN integrated sequence analysis (RISA) system--384-format  
**JOURNAL** sequencing pipeline with 384 multicapillary sequencer  
**MEDLINE** Genome Res. 10 (11), 1757-1771 (2000)  
**PUBMED** 20530913  
**REFERENCE** 11076861  
**AUTHORS** 4  
 The RIKEN Genome Exploration Research Group Phase II Team and the  
 FANTOM Consortium.  
**TITLE** Functional annotation of a full-length mouse cDNA collection  
**JOURNAL** Nature 409, 685-690 (2001)  
**REFERENCE** 5  
 The FANTOM Consortium and the RIKEN Genome Exploration Research  
 Group Phase I & II Team.  
**AUTHORS** Analysis of the mouse transcriptome based on functional annotation  
 of 60,770 full-length cDNAs  
**JOURNAL** Nature 420, 563-573 (2002)  
**REFERENCE** 6 (bases 1 to 2373)  
**AUTHORS** Aachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carlnici, P.,  
 Fukuda, S., Furuno, M., Hanagaki, T., Har, A., Hashizume, W.,  
 Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T.,  
 Horii, P., Imocani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T.,  
 Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M.,  
 Koye, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M.,  
 Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N.,  
 Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N.,  
 Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T.,  
 Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S.,  
 Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A.,  
 Muramatsu, M. and Hayashizaki, Y.  
**TITLE** Direct Submission  
**JOURNAL** Submitted (16-JUL-2001) Yoshihide Hayashizaki. The Institute of  
 Physical and Chemical Research (RIKEN), Laboratory for Genome  
 Exploration and Chemical Research (RIKEN), Laboratory for Genome  
 Exploration Research Group, RIKEN Genomic Sciences Center (GSC),  
 RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,  
 Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.jp,  
 URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222,  
 Fax: 81-45-503-9216)  
**COMMENT** CDNA library was prepared and sequenced in Mouse Genome  
 Encyclopedia Project of Genome Exploration Research Group in Riken  
 Genomic Sciences Center and Genome Science Laboratory in Riken.  
 Division of Experimental Animal Research in Riken contributed to  
 prepare mouse tissues.  
 Please visit our web site for further details.  
 URL: http://genome.gsc.riken.jp/  
 Location/Qualifiers  
**FEATURES**  
 source  
 1. .2373  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="C57BL/6J"  
 /db\_xref="FANTOM\_DB:9430031M11"  
 /db\_xref="taxon:10090"  
 /clone="9430031M11"  
 /tissue\_type="embryonic body between diaphragm region and  
 neck"  
 /clone\_1lb="RIKEN full-length enriched mouse cDNA library"  
 /dev\_stage="12 days embryo"  
 43. .2011  
 /note="Ewing sarcoma homolog (MGI:99960, GB|NM\_007968,  
 evidence: BLASTN, 99%, match=2172)  
 putative"  
 2358. .2363  
 /note="putative"  
 2373  
 /note="putative"  
**ORIGIN**  
 Alignment Scores:  
 Pred. No.: 2,64e-206 Length: 2373  
 Score: 3530.50 Matches: 638  
 Percent Similarity: 98.48% Conserves: 9  
 Best Local Similarity: 97.11% Mismatches: 8  
 Query Match: 97.18% Indels: 2

DB: 3 Gaps: 1  
 US-10-791-017a-2 (1-656) x AK034755 (1-2373)  
 QY 1 MetAlaSerThrApPTrYSerThrTrYSerGlnAlaAlaGlnGlnGlyTrYSerAla 20  
 Db 43 ATGGCGTCCAGGATTTACGATCTTATGATCAAGCTCCAGCGGGCTACAGTCT 102  
 QY 21 TTTThrAlaGlnProThrglnGlyTTrAlaGlnThrGln-AlaTTrGlnGlnSe 40  
 Db 103 TACACGCCCGCAGCACTCAAGATATGACAGACGCCAGGATATATGGCAACAAG 162  
 QY 40 TTrYGlyTrThrTTrYGlyGlnProThraPpAlaSerTTrThrGlnAlaGlnThrAlaTh 60  
 Db 163 CTATGAACTTATGACAGGCTCTAGTGTGACCTATATCAGGCTCAGACCACTGCGAC 222  
 QY 60 TTrYGlyGlnThrAlaTTrAlaThraPpAlaSerTTrYGlyGlnProThrGlnGlyTTrThrPr 80  
 Db 223 CTACGGGCAAGCTGATATGCACTTCTTACGACGCTCCAGCTGTTATGACCTCC 282  
 QY 80 oThraAlaProGlnAlaTTrYSerGlnProValGlnGlyTTrGlyAlaTTrAlaPpTh 100  
 Db 283 AACGCGCCCGCAGCGGTAACAGCCAGCTGTGAGGGATATGGACCTGAGCTTATGACAG 342  
 QY 100 TTrThrAlaThraValThrThrThrGlnAlaSerTTrAlaAlaGlnSerAlaTTrGlyTh 120  
 Db 343 CACCACTGCTACAGTCAACCAAGCGGCGCTTACGAGCTCAGCTCAGCTCATATGGCAC 402  
 QY 120 TTrAlaProAlaTTrProAlaTTrYGlyGlnGlnProAlaAlaThraAlaProThraPpProG 140  
 Db 403 CAGGCTGCTTACCTCCACCTATATGCGCAGCGCAAGCCAGCCACTTACAGACACA 462  
 QY 140 naPpGlyAsnLysPProThrglnThrThrSerGlnProGlnSerThrglyTTrYaaGln 160  
 Db 463 GGAATGGTAACAAGCTCTGCTAGCTAGTCAACTCAATATGACAGGGGTTATATACCA 522  
 QY 160 nProSerLeuGlyTTrYGlyGlnSerAntYTrYSerTTrProGlnValPProGlySerTTrPr 180  
 Db 523 ACCAGCGTATGAGTATGACAGAGTAACTACACTATCCCACTATCCAGTCTGAGCTACCC 582  
 QY 180 oMetGlnProValThraAlaProProSerTTrProProThrSerTTrYSerSerThrglnPr 200  
 Db 583 AATCGACCGCATGACCGCACTTACTTCTCTTACAGCTTCTCTTCAACGCC 642  
 QY 200 oThrSerTTrAspGlnSerSerTTrYSerGlnGlnAlaThrTTrYGlyGlnProSerSerTTr 220  
 Db 643 GACTAGTTACGATCAGAGGAGTTACTCTACAGAACTATAGGGAGCGAGCGACTA 702  
 QY 220 TTrYGlnGlnSerSerTTrYGlyGlnGlnSerSerTTrYGlyGlnGlnProProThrSerTTr 240  
 Db 703 TGGACAAACAGAGTATGCTATGTCACAAAGCACTATGGGACAGACCTCTCTATGTTA 762  
 QY 240 rPProProGlnThrGlnSerTTrYSerGlnAlaProSerGlnTTrYSerGlnGlnSerSerSe 260  
 Db 763 CCCCCCTCAGACTCGATCTTACAGCGAGGCTCCAACTCAATATAGCCACAGACAGCAG 822  
 QY 260 TTrYGlyGlnGlnSerSerThraGlnAlaPpAlaProSerSerMetGlyValTTrGlyGln 280  
 Db 823 CTACGGGCAAGAGTTCATCTCCGACAGAGCAACCCCAAGTATGAGTATATGGCA 882  
 QY 280 nGlnSerGlyGlyPhaSerGlyProGlyGlnAlaAlaGSerMetSerGlyProAlaPpAla 300  
 Db 883 GAGGTCTGAGAGATTTTCGAGACAGAGAGAAACCGAGCTTGAATGGCCCTCATTAACG 942  
 QY 300 GGIYAARGIYAARGIYGIYPheAPARGLIYMETSERARGLIYGIYAARGIYGIYGI 320  
 Db 943 GGGCAGGAGAGAGGGGATTTATCTGAGGCAAGACAGAGGTGGGGGAGAGAGG 1002  
 QY 320 YARGIYGIYMETGlySerAlaGlyGlnuArgIYGIYPheAsnLysPProGlyPProWe 340  
 Db 1003 ACGCGGTGAGCTGGG---GCTGAGAGAGAGGTGCTTCAATTAACCTGTGTGACCAT 1059  
 QY 340 tAepGIuGIYProAlaPpLeuAlaPpLeuGIYProProValaPpProAlaPpSerAlaPp 360



```

Db      1060 GAGTGAAGACAGATTCTTATCTAGAGCTTCCATATGATCCGATGAAGACTCTGCAAA 1119
Qy      360 nserAlIleTyValnIngIyLeuAnaBpSerValThIleuAspApleuAlaBpH 380
Db      1120 CAGGCAATTTATCTGACAGATTAAATGACAAATGATCACTGATGATCTGGACAGACTT 1179
Qy      380 epehlyeGhIcyegIyValIyValIyBmeCtAmIyBArXThGhIyGhInPmeCtIleHsII 400
Db      1180 CTTTAAACAGTGTGGGGTGTCTCAAGATGAACAGAGAACCTGAGAACACCTATGATCATAT 1239
Qy      400 eTyTleuAspLyegIuThrGhIyTyPProLyegIyAspAlaThrValSerTyTgIuAspPr 420
Db      1240 CTACCTGATTAAGAGACAGAGAAACCTTAAAGGGGAGCCACAGTGTCTATGAAGATCC 1299
Qy      420 oPrOThAlAlYAlAlAlAlValGtUTTPheAspGhIyLyAspPhelngIySerIyBle 440
Db      1300 ACCAACTGCAAAAGGCTCCCTGATGATGATGGGAAAGATTTCACAGAGCAAAACT 1359
Qy      440 uLySValSerIleuAlaArgLySlySPrOPrometAsnSerMetArgLyGhIyLeuPProP 460
Db      1360 TTAAGTCTCTTCCCGGAAAGAGCTTCCATTAAGACATGCGGGAGGCAATGCCACC 1419
Qy      460 oATgIuGhIyArgLyMeCProPProPoleuArgLyGhIyProGhIyPProGhIyPPr 480
Db      1420 TCGTGAAGGAGGAGGAGATGCACACCACTTCGTGAGGTCTGTGGCCAGAGAGGCC 1479
Qy      480 oGhIyGhIyPProMeGhIyArgMetGhIyGhIyArgLyGhIyAspArgLyGhIyPhePProAr 500
Db      1480 TGAAGAGCCATGGGTGCAATGGAGGCGTGGAGGACCAAGAGGAGGCTTCCCTCCAG 1539
Qy      500 gGhIyProArgGhIySerArgLyAsnProSerGhIyGhIyAsnValGhIyAsnValGhIy 520
Db      1540 AGGGCCCCGAGGCTCCAGAGAGAAACCTCTGAGAGAGAAATGTCAGACACGAGCTGG 1599
Qy      520 yAsPTrpGhIyCyBProAsnProGhIyCyGhIyAsnGhIyAsnPhelAlaTrpArgThrgIuCy 540
Db      1600 AGACTGGCAGTGTCCCAATCAAGGCTGTGMAACCAAGACTTGTGGAAGACAGATG 1659
Qy      540 BAENGInCyLyAlAlArProLyBProGhIyGhIyPheLeuPProPProPhePProPProGhI 560
Db      1660 CAACCAAGTAAAGCCCTTAAGCCCGAGGCTTCTCCGCAACCTTTCACCTCCGGG 1719
Qy      560 yGhIyAspArgGhIyArgLyGhIyProGhIyGhIyMetArgLyGhIyArgLyGhIyLeuMetAs 580
Db      1720 TGGTGTATCGTGAAGAGAGTGGCTCTGTGGCATGCGAGAGAGAGAGAGAGAGAGAGAGAG 1779
Qy      580 pArgGhIyGhIyProGhIyGhIyMetPheArgLyGhIyArgLyGhIyAspArgLyGhIyPheAr 600
Db      1780 CCGTGTGTCTCTGAGAGATGTTCAAGAGTGGGAGAGTGGAGACAGAGAGAGGCTTCCG 1839
Qy      600 gGhIyGhIyArgLyGhIyMetAspArgLyGhIyPheGhIyGhIyArgLyGhIyProGhIyGhI 620
Db      1840 AGGTGGCCGTGAGATGAACGAGAGTGGCTTGTGAGAGAGAGAGAGAGAGAGAGAGAGAG 1899
Qy      620 yProProGhIyProleuMetGhIyGhIyMetGhIyGhIyArgArgLyGhIyArgLyGhIyProGhI 640
Db      1900 GCTCTCCGAGACTTTATGACAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1959
Qy      640 yLySmetAspLyegIyGhIyAsnArgIngIuArgArgAspArgProTyR 656
Db      1960 GAAATATGATTAAGGAGACACCTCAGAACGAGAGAGAGAGAGAGAGAGAGAGAGAG 2008

```

```

RESULT 12
LOCUS   CR604639
DEFINITION full-length cDNA clone CS0DJ013YN08 of T cells (Jurkat cell line)
ACCESSION CR604639
VERSION   CR604639.1 GI:50485446
KEYWORDS  HTc; CNSLT; cDNA.
SOURCE    Homo sapiens (human)
ORGANISM  Homo sapiens

```

```

REFERENCE
AUTHORS  Li, W.-B., Gruber, C., Jessee, J. and Polyes, D.
TITLE     Full-length cDNA libraries and normalization
JOURNAL   Unpublished
REMARK    Contact : Feng Liang Email : fliang@lifetech.com URL :
          http://fulllength.invitrogen.com/InvitrogenCorporation1600
          Parady Avenue
          Genoscope.
          2 (bases 1 to 2236)

REFERENCE
AUTHORS  Direct Submision
TITLE     Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage :
JOURNAL   BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
          - Web : www.genoscope.cns.fr)

COMMENT   1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
          end enriched, double-strand cDNA was digested with Not I and cloned
          into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
          was normalized. Library was constructed by Life Technologies, a
          division of Invitrogen.

FEATURES
Source    Location/Qualifiers
           1..2236
           /organism="Homo sapiens"
           /mol_type="mRNA"
           /db_xref="taxon:9606"
           /clone="CS0DJ013YN08"
           /release_type="T cells (Jurkat cell line) cot
           10-normalized"
           /plasmid="pCMVSPORT_6"

ORIGIN
Alignment Scores:
Pred. No.:      9 74e-202      Length:      2236
Score:          3455.50      Matches:      632
Percent Similarity: 96.34%      Conservative: 0
Best Local Similarity: 96.34%      Mismatches: 1
Query Match:    95.11%      Indels:      24
DB:              3          Gaps:          2

```

US-10-791-017a-2 (1-656) x CR604639 (1-2236)

```

Qy      1 MetAlSerThrAspTySerThrTySerGlnAlAlAlGlnGlnGlnTySerAla 20
Db      19 ATGGCTTCCACCGATTACGATCTATACCAAGCTGACGAGCGGAGCTTCAAGTCT 78
Qy      21 TyTThAlGlnProThrGlnGlnTyTAlaGlnThThGlnAlAlAlTyGlnGlnSer 40
Db      79 TACACCGCCACGCGCACTCAAGATATGACACAGACCAAGCAGCATATGGGCAAAAGC 138
Qy      41 TyTGTThTyTGTGlnProThraBpValSerTyTThGlnAlaGlnThThAlaThr 60
Db      139 TATGAACTTATGACAGAGCCCACTATATGACATTAACCAAGCTAGACACTGCAAC 198
Qy      61 TyGhIyGlnThAlaTyAlaThrSerTyGhIyGlnProProThrgIyTyThThPro 80
Db      199 TATGGCAGAGCGGCTATCACTTATATGACAGAGCTCCCACTGTTATACTCCCA 258
Qy      81 ThAlaPProGlnAlAlTySerGlnProValGlnGlnTyTGTThTyTAlaTyAlaThr 100
Db      259 ACTGCCCCCAGACATACAGCGAGCTGTCCAGGGGTATGGCATGTGCTTATGATACC 318
Qy      101 ThThAlaThrValThThThThGlnAlaSerTyTAlaAlaGlnSerAlaTyTGTThr 120
Db      319 ACCACTGCTACAGTACCAACCAAGGCTCTCTATGACGCTCAGTCTGATATGGCACT 378
Qy      121 GlnProAlaTyProAlaTyTGTGlnGlnProAlaAlaThAlaBProThraBProGln 140
Db      379 CAGCGCTTATTCAGGCTATGGGAGAGAGCAGCAGCAGCACTGACCTTACAGACGGAG 438
Qy      141 AspGhIyAsnLyBProThrgIuThrSerGlnProGlnSerSerThrgIyGhIyTAAngIn 160
Db      439 GATGAAACAAAGCCCACTGAGACTATGTAACCTCAATCTTACAGACAGAGGGGTAAACAG 498

```



QY 161 ProSerLeuGlyTyrGlyGlnInsertAsnTyrSerTyrProGlnValProGlySerTyrPro 180  
 DB 499 CCAGGCTGAGATATGACAGAGATTAACATGATATCCAGGATACCTGGAGACTACCCC 558  
 QY 181 MetGlnProValThrAlaProSerTyrProProthSerTyrSerSerThrGlnPro 200  
 DB 559 ATGCACCACTGACTGACCTCCATCTCACTCTCAACGATATTCCTTCAACAGCCG 618  
 QY 201 ThrSerTyrAspGlnSerSerTyrSerGlnGlnAsnThrTyrGlyGlnProSerSerTyr 220  
 DB 619 ACTAGTATATGATCAGAGACAGTACTCTCAGCAGAAACCTATGGGACCGACAGCTAT 678  
 QY 221 GlyGlnGlnSerSerTyrGlyGlnGlnInsertSerTyrGlyGlnGlnProProthSerTyr 240  
 DB 679 GGCACACAGATGCTATGCTCAACAAAGCAGTATGGGACGAGCTCCACTAGTATAC 738  
 QY 241 ProProGlnThrGlySerTyrSerGlnAlaProSerGlnTyrSerGlnGlnInsertSer 260  
 DB 739 CCACCCCAACTGGATCTCAACCCAGCTCCAGTCAATATGACCAACAGACAGACAGC 798  
 QY 261 TyrGlyGlnGlnSerSerPheArgGlnAspHisProSerSerMetGlyValTyrGlyGln 280  
 DB 799 TACGGGACAGAGATTCATCCGACAGAACCCAGTAGCATGGGTGTTATGGGACG 858  
 QY 281 GluSerGlyGlyPheSerGlyProGlyGlyLeuAsnArgSerMetSerGlyProAspAsnArg 300  
 DB 859 GAGTCTGAGAGATTTTCCGACAGAGAACCGAGCATGTGGCCCTGATTAACCGG 918  
 QY 301 GlyArgGlyArgGlyGlyPheAspArgGlyGlyMetSerArgGlyGlyArgGlyGlyGly 320  
 DB 919 GCGAGGGGAAAGAGGGGATTTGATCTGAGGCAATGACGAGAGTGGCCGGAGAGAGA 978  
 QY 321 ArgGlyGlyMetGlySerAlaGlyGlyGlyGlyPheAsnArgProGlyGlyProMet 340  
 DB 979 CCGGGTGGATGGGC---GCTGGAGAGCAGGTGGCTTCAATAGCTGGTGGACCCATG 1035  
 QY 341 AspGlyGlyProAspLeuAspLeuGlyProProValAspProAspGlyLeuAspSerAspAsn 360  
 DB 1036 GATGAAGGAGACCAATCTGATCTAGGCCCCAGTGAATCAATGAACTCTGACAAAC 1095  
 QY 361 SerAlaIleTyrValGlnGlyLeuAsnAspSerValThrLeuAspAspLeuAlaAspPhe 380  
 DB 1096 AGTGCATTTTATTAAGATTAATGACAGTGTACTTAATGATCTGGAGACTTC 1155  
 QY 381 PheGlyGlnGlyGlyValValValMetAsnLysArgThrGlyGlnProMetIleHisIle 400  
 DB 1156 TTTTAACAGAGTGGGGTGTGTAAGATGAACAAGAACTGGGCAACCATGATCCACATC 1215  
 QY 401 TyrLeuAspLysGlyLeuThrGlyLysProLysGlyAspAlaThrValSerTyrGlyLeuAspPro 420  
 DB 1216 TACCTGACAAAGAAACAGGAAAGCCCAAGGAGATGCCACAGTGTCTTATGAAGACCCCA 1275  
 QY 421 ProThrAlaValAlaValAlaValGluTTPPheAspGlyLysAspPheGlnGlySerLysLeu 440  
 DB 1276 CCACCTGCCCAAGGCTCCGCTGGAATGTTGATGGGAAGATTTTCAAGGAGCAAACTT 1335  
 QY 441 LysValSerLeuAlaArgLysLysProProMetAsnSerMetArgGlyGlyLeuProPro 460  
 DB 1336 AAAGTCTCCCTGCTCGAAGAAAGCCCTCAATGAACGATATGGGGGTGTCTGCCACCC 1395  
 QY 461 ArgGlyGlyArgGlyMetProProProLeuAspGlyGlyProGlyGlyProGlyGlyPro 480  
 DB 1396 CCGAGAGGAGAGAGGATGCCACCACTCCGAGAGGTCCAGAGAGCCCAAGAGGTCTCT 1455  
 QY 481 GlyGlyProMetGlyArgMetGlyArgGlyGlyAspArgGlyGlyPheProProArg 500  
 DB 1456 GGGGAGACCATGGGTGCGATGGAGGCGCTGAGAGAGATGAGAGGCTTCCCTCCAAAG 1515  
 QY 501 GlyProArgGlySerArgGlyAsnProSerGlyGlyGlyAsnValGlnHisArgAlaGly 520  
 DB 1516 GGAACCCCGGGGTTCGAGAGGAAACCCCTCTGAGAGAGAAACGTCCAGACCGAGCTGGA 1575  
 QY 521 AspTyrGlnCysProAsnProGlyCysGlyAsnGlnAsnPheAlaTrrpaGthrGlyCys 540

DB 1576 GACTGGACAGTGTCCCAATCCGGGTGTGGAAACCAAACTTCCCTCGAGAACAGAGTGC 1635  
 QY 541 AsnGlnCysLeuAlaProLysProGlyGlyPheLeuProProProPheProProGly 560  
 DB 1636 AACCACTGTAAGGCCCCCAAGCCTGAAGGCTTCTCCGCAACCTTTCGGCCCGGGGT 1695  
 QY 561 GlyAspArgGlyArgGlyGlyProGlyGlyMetArgGlyGlyArgGlyGlyLeuMetAsp 580  
 DB 1696 GGTGATCGTGGCAGAGGTGGCCCTGTGGCATGCGGGGAGAAAGAGTGGCTTCATGAT 1755  
 QY 581 ArgGlyGlyProGlyGlyMetPheArgGlyGlyArgGlyGlyAspArgGlyGlyPheArg 600  
 DB 1756 CGTGTGTTCCTCGGTGAATGTTCAAGAGTGGCCGTGTGGAGACAGAGTGGCTTCGT 1815  
 QY 601 GlyGlyArgGlyMetAspArgGlyGlyPheGlyGlyGlyArgGlyGlyProGlyGly 620  
 DB 1816 GTGTGCGGGGAGATGACCGAGGTGGCTTGTGTGAGAGAAACGAGTGGCCCTGGGGGG 1875  
 QY 621 ProProGlyProLeuMetGlnMetGlyArgArgGlyGlyArgGlyGlyProGly 640  
 DB 1876 CC-CCA----- 1880  
 QY 641 LysMetAspLysGlyGlyHisArgGlnGlyArgArgAspArgProTyr 656  
 DB 1881 -----GGCAGACCGCTCAGAGCGCAGAGATCGCCCTTAC 1916  
 RESULT 13  
 AK049743  
 LOCUS AK049743  
 DEFINITION Mus musculus 12 days embryo spinal cord cDNA, RIKEN full-length enriched library, clone: C530046A18 product: Ewing sarcoma homolog, full insert sequence.  
 ACCESSION AK049743  
 VERSION AK049743.1 GI:26093614  
 KEYWORDS HTC; CAP trapper.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 REFERENCE  
 AUTHORS Carninci, P. and Hayashizaki, Y.  
 TITLE High-efficiency full-length cDNA cloning  
 JOURNAL Meth. Enzymol. 303, 19-44 (1999)  
 MEDLINE 99279253  
 PUBMED 10349636  
 REFERENCE  
 AUTHORS Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Komoto, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
 TITLE Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes  
 JOURNAL Genome Res. 10 (10), 1617-1630 (2000)  
 MEDLINE 20499374  
 PUBMED 11042159  
 REFERENCE  
 AUTHORS Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Komoto, H., Akiyama, J., Nishi, K., Kitsuai, T., Taahito, H., Itoh, M., Sumi, N., Iehi, Y., Nakamura, S., Hazama, M., Nishino, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwake, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Matshiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, O., Okazaki, Y., Muramatsu, M., Inoue, Y., Kita, A. and Hayashizaki, Y.  
 TITLE RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multichannel sequencer  
 JOURNAL Genome Res. 10 (11), 1757-1771 (2000)  
 MEDLINE 20510913  
 PUBMED 11076861  
 REFERENCE  
 AUTHORS The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.  
 TITLE Functional annotation of a full-length mouse cDNA collection  
 JOURNAL Nature 409, 695-690 (2001)  
 REFERENCE 5

**AUTHORS** The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.  
**TITLE** Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs  
**JOURNAL** Nature 420, 563-573 (2002)  
**REFERENCE** (bases 1 to 2269)  
**AUTHORS** Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hirooka, T., Hirozane, T., Horii, P., Imocant, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Kato, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsumura, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Ozaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Mahira, S., Takekida, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M., and Hayashizaki, Y.

**TITLE** Direct Submission  
**JOURNAL** Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gsr.riken.jp, URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)

**COMMENT** cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in Riken. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.  
 Please visit our web site for further details.  
 URL: http://genome.gsc.riken.jp/  
 URL: http://fantom.gsc.riken.jp/  
 Location/Qualifiers

**FEATURES****source**

1. 2269  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="C57BL/6J"  
 /db\_xref="FANTOM:DB:C530046A18"  
 /db\_xref="taxon:10090"  
 /clone="C530046A18"  
 /tissue\_type="spinal cord"  
 /issue\_type="RIKEN full-length enriched mouse cDNA library"  
 /dev\_stage="12 days embryo"  
 48..1903  
 /note="Ewing sarcoma homolog (MED)|MGI:99960, GB|NM\_007968, evidence: BLASTN, 99%, match=2172)  
 putative"  
 2251..2256  
 /note="putative"  
 2269  
 /note="putative"  
 /note="putative"

**ORIGIN****Alignment Scores:**

Pred. No.: 1,74e-193 Length: 2269  
 Score: 3321.00 Matches: 606  
 Percent Similarity: 93.45% Conservative: 7  
 Beel Local Similarity: 92.38% Mismatches: 5  
 Query Match: 91.41% Indels: 39  
 DB: 3 Gaps: 1

US-10-791-017a-2 (1-656) x AK049743 (1-2269)

QY 1 MetAlserThraplyrSerThrTYrSerGlnAlaAlaAlaGlnGlnGlnTYrSerAla 20  
 DB 48 ATGGCGCTCCCGGATTAAGTACCTAATGCAAGCTCAGCCAGCGGCTCAGTGT 107  
 QY 21 TYrThAlaGlnProThrGlnGlnTYrAlaGlnThrThrGlnAlaTYrGlnGlnGlnSer 40  
 DB 108 TACACGCCGCAACTCAAGGATATGACAGACCAAGGACATATGGGCAACAAAGC 167

QY 41 TYrGlyThrTYrGlyGlnProThrAapValSerTYrThrGlnAlaGlnThrThrAlaThr 60  
 DB 168 TATGAACCTATGAGACGCTACTATGTACGTAATCTACGCTACAGCACTGCCACC 227  
 QY 61 TYrGlyGlnThrAlaTYrAlaThrSerTYrGlyGlnProProThrGlyTYrThrThrPro 80  
 DB 228 TACGGGACAGTCGATACCACTTCTTACGAGACCTCCCACTGGTATATACACTCCA 287  
 QY 81 ThrAlaProGlnAlaTYrSerGlnProValGlnGlnTYrGlyThrGlyAlaTYrAapThr 100  
 DB 288 ACTGCCCCGAGCGCTGACGACGAGCTGTGAGGGATATGGCACTGGGACTTATGACAGC 347  
 QY 101 ThrThAlaThrValThrThrThrGlnAlaSerTYrAlaAlaGlnSerAlaTYrGlyThr 120  
 DB 348 ACCACGCTACAGTCAACCAACGAGGCTCTTACGAGCTACGACTGATATGGCACC 407  
 QY 121 GlnProAlaTYrProAlaTYrGlyGlnGlnProAlaAlaThrAlaProThrAapProGln 140  
 DB 408 CAGCTGCTACCCCTACCTATGGCCAGACCAAGCCAGCCGCACTACAGACCAAG 467  
 QY 141 AapGlyAapThyProThrGlnThrSerGlnProGlnSerSerThrGlyGlyTYrAapGln 160  
 DB 468 GATGTAAACAGCTGCTGAGTCACTCACTCAATCTAGCAAGAGGGGTATTAACCA 527  
 QY 161 ProSerLeuGlyTYrGlyGlnSerAanTYrSerTYrProGlnValProGlySerTYrPro 180  
 DB 528 CCAAGCTGATGATATGAGACAGTAACTACAGTATCCCAAGTACTGGAGCTACCA 587  
 QY 181 MetGlnProValThrAlaProProSerTYrProProThrSerTYrSerSerThrGlnPro 200  
 DB 588 ATGCACCGCATCACCACCTCCATCTTATCCCTACAGCACTACCTTCCACAGCCG 647  
 QY 201 ThrSerTYrAapGlnSerSerTYrSerGlnGlnAanThrTYrGlyGlnProSerSerTYr 220  
 DB 648 ACTAGTACGATCAAGCACTTACTCTGACGAAACCTATAGGGAGCCGACACTAT 707  
 QY 221 GlyGlnGlnSerSerTYrGlyGlnGlnSerSerTYrGlyGlnGlnProProThrSerTYr 240  
 DB 708 GAGCAACAGAGTACGATATGTGTCAACAAACAGCTATGGGACAGAGCTCTACTATTC 767  
 QY 241 ProProGlnThrGlySerTYrSerGlnAlaProSerGlnTYrSerGlnGlnSerSerSer 260  
 DB 768 CCCCCTCAAGCTGATCTTCAAGCCAGCTCCCAAGTCAATATAGCCAAAGAGCAGCAGC 827  
 QY 261 TYrGlyGlnGlnSerSerPheAapGlnAapPheAapProSerSerMetGlyValTYrGlyGln 280  
 DB 828 TACGGGACAGAGTTCATTCGACAGGACCAACCCAGATGACATGGGTATATGGGACG 887  
 QY 281 GluSerGlyGlyPheSerGlyProGlyGlnAanArgSerMetSerGlyProAapAanArg 300  
 DB 888 GAGTCTGAGGATTTTCCGAGCCAGAGAGAAACCGGAGCTTGAATGGCCCTGATTAACCG 947  
 QY 301 GlyAargGlyAargGlyGlyPheAapAargGlyGlyMetSerAargGlyGlyAargGlyGly 320  
 DB 948 GGCAGGGGAAGAGGGGATTTGATCTGAGGACATAGCAGA----- 989  
 QY 321 ArgGlyGlyMetGlySerAlaGlyAlaGlyGlyPheAanLysProGlyGlyPromet 340  
 DB 989 ----- 989  
 QY 341 AapGlyGlyProAapLysAapLysGlyProProValAapProAapGlyLysAapAan 360  
 DB 990 -----GATCCGATGAGACCTGACAAAC 1013  
 QY 361 SerAlaIleTYrValGlnGlyLeuAanAapSerValThrLeuAapLysLeuAlaAapPhe 380  
 DB 1014 AGTGCATTTATGTCGCAAGATTAATGCAATATGATCTGATATCTGCGACGACTTC 1073  
 QY 381 PheLysGlnCysGlyValValLysMetAanLysAargThyGlyGlnPrometIleHsIle 400  
 DB 1074 TTTAAGCACTGTGGGTGTTCAGATGAACAGAGAACTGACCAACCCATGATCCATATC 1133  
 QY 401 TYrLeuAapLysGlyThrGlyLysProLysGlyAapAlaThrValSerTYrGlyAapPro 420



QY 139 ProGlnAspGlyValAsnLysProThrGlnThrSerGlnProGlnSerThrGlyTyr 158  
 Db 431 CCTCAAGATGGAAGCAAGCAAGATACCAATAGCCCAACCAAGTACAGCACTAT 490  
 QY 159 AsnGlnProSerLeuGlyTyrGlyGlnSerAsnTyrSerTyrProGlnValProGlySer 178  
 Db 491 GCCAGGCTGACATGGCTATAGTCAAAAGACAGTACAGTACCTCAAGTGTGTCAGC 550  
 QY 179 TyrProMetGlnProValThrAlaProProSerTyrProProThrSerTyrSerSerThr 198  
 Db 551 TATCCAGTGAAGCAAGTGAAGGACCAACATCTTACCTCTCTCA-----AGCTCC 601  
 QY 199 GlnProThrSerTyrAspGlnSer----- 206  
 Db 602 CAGCATCAAGCATATGAACAGAGCTCTATCCAGCCAAAGTCTTACACCAAGAGAGC 661  
 QY 207 -----SerTyrSerGlnGlnAsnThrTyrGlyGlnProSerSerTyr 220  
 Db 662 AGCTATGACAGCAAACTTCATATTCACAGAGGCTCTTATGACAGCAAGTAT 721  
 QY 221 GlyGlnGlnSerSerTyrGlyGlnGlnSerSerTyr-----GlyGlnGln 235  
 Db 722 GGTCAACAAAGTGAATATGACCAAGATAGTCAAGCAAGCAAGCAAGCAAGCAAGT 761  
 QY 236 ProProThrSerTyrProProGlnThrGlySerTyr-----SerGlnAlaProSerGlnTyr 254  
 Db 782 CTTCTTACCAAGTATACCACTCTTATCTTCTATAGCAATGACCAAGCAAGCAAGTAT 841  
 QY 255 SerGlnGlnSerSerSerTyrGlyGlnGlnSerSerPheArgGlnAspHisProSerSer 274  
 Db 842 GGGCAACAGAGCAAGTATGACCAAGCAAGCAAGCAAGTATGTCAGCAAACTTCAAGAGT 901  
 QY 275 MetGlyValTyrGlyGlnGln-----SerGlyGlyPheSerGlyProGlyGlnAsnArgSer 293  
 Db 902 AAAAGTCCCTATGACCAAGGCTCTATCAGAGGTTTCTCCAGCTCTGCAATATCTGAGT 961  
 QY 294 MetSerGlyProAspAsnArgGlyValArgGlyValGlyPheAspArgGlyGlyMetSer 313  
 Db 962 TCTGAGAGCTCCGATTCACGACAGAGGCTCGGGGATGTTTGAAGAGTGGCATAGC 1021  
 QY 314 ArgGlyGlyValArgGlyGlyValArgGlyGlyMetGlySerLeuGlyGlnArgGlyGlyPhe 333  
 Db 1022 CGTGGTGGAGG---GGCTCTCGTGGGTCATGGC---GGTGAAGATCAGAGTGGCTTC 1075  
 QY 334 AsnLysProGlyValProMetAspGlnGlyProAspLeuAsnLeuGlyPro----- 350  
 Db 1076 AGTAACTGCTGATGATCTTGGATGATGACAGATCTTATTTAGGACCCCTATAGCT 1135  
 QY 351 -----ProValAspProAspGlnAspSerAspAsnSerAlaLeuTyrValGlnGly 367  
 Db 1136 TTACCATTTGCTTGGACCCAGAAAGAGCAAGAAACAAATCTGTATGTACAGGGA 1195  
 QY 368 LeuAsnAspSerValThrLeuAspAspLeuAlaAspPheLeuGlnCysGlyValVal 387  
 Db 1196 CTAAAGAGATTAATCTGTGAAGAAATGTGATTTCTTCAACACCTGTGTGATGTC 1255  
 QY 388 LysMetAsnLysArgThrGlyGlnProMetIleHisLeuLeuAspLysGlyGlnThrGly 407  
 Db 1256 AGATTAACAGAGGACAGGCAACCACTTTAACTTTTACGACAGAGAAACCGG 1315  
 QY 408 LysProGlyGlyAspAlaThrValSerTyrGlnAspProProThrAlaLysAlaVal 427  
 Db 1316 AAGCCCAAGAGATGACATGCTGTCTTTGAAGACCACTTCACTAAGATGCTGTA 1375  
 QY 428 GlnTyrPheAspGlyLysAspPheGlnGlySerLeuLeuValSerLeuAlaGlyLys 447  
 Db 1376 GAACCTTGTATGGAGAGATCTTAATGAGAGCAAGTGAAGTGTGCTGCTCCCAA 1435  
 QY 448 LysProProMetAsnSerMetArgGlyGlyLeuProProArgGlyGlnGlyArgMetPro 467  
 Db 1436 AAGCTTTGCTAGTACATGACAGAGGGGCTTCTTTTACGTACAGCAAGAGGAGCCCA 1495  
 QY 468 ProProLeuArgGlyGlyProGlyGlyProGlyGlyProGlyGlyProMetGlyArgMet 487

Db 1496 CCACCTCTCCGT-----GGAGCTCCCAAGGGTCCGATG 1528  
 QY 488 GlyGlyValArgGlyValAspArgGlyGlyPheProProArgGlyProArgGlySerArgGly 507  
 Db 1529 GGTGGCCGAGAGAGAACCGTGGGATTCATGCCAAAGAGACACAGTGAAGCTCTTAA 1588  
 QY 508 AspProSerGlyGlyValAsnValGlnHisArgAlaGlyAspTyrGlnCysProAspPro 527  
 Db 1589 AGCCCAAGT---AGCGTATATGTCAGCACAGGAGCTGTGATCGCATGTCCCAACCG 1645  
 QY 528 GlyGlyValAsnGlnAsnPheAlaTyrPheGlnGlnCysAsnGlnCysValAlaProLys 547  
 Db 1646 GATGTGAACCAAAATTTTGTCTGAGAACAGATATGTAACAGTGAAGCTCTTAA 1705  
 QY 548 ProGlnGlyPheLeuProProPheProProProGlyGlyValAspArgGlyValArgGly 567  
 Db 1706 CCGTATGCTCTTTT---CCTCTTCCCACTCAAGAGAGAACAGTGAAGAGTGG 1762  
 QY 568 ProGlyGlyMetArgGlyValArgGlyGlyLeuMetAspArgGlyGlyProGlyGlyMet 587  
 Db 1763 CCC---GGCATGAGAGAGGACAG---GGCTTATGACCGTGAAGGCT---GGTATG 1813  
 QY 588 PheArgGlyValArgGlyValAspArgGlyGlyPheArgGlyGlyValArgGlyMetAspArg 607  
 Db 1814 TTCAGAGGTGGCCGTGAGAGATAGAGGGGGCTTCA---GGCGTGGCATGACAGA 1870  
 QY 608 GlyGlyPheGlyGlyValArgArgGlyGlyProGlyGlyProProGlyProLeuMetGln 627  
 Db 1871 ---GATTTGTGAG 1927  
 QY 628 GlnMetGlyValArgArgGlyGly-----ArgGlyGlyProGlyLysMetAspLysGly 645  
 Db 1928 CAATCTGTGGAAGAGTGTGTGTGAAGACAGAGTGTCTCCAGCAAAATGATTAAGT 1987  
 QY 646 GlnHisArgGlnGlnArgArgAspArgProTyr 656  
 Db 1988 GACCATGCGCAGAGAACGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2020  
 RESULT 15  
 AK014366 1379 bp mRNA linear HTC 03-APR-2004  
 LOCUS AK014366  
 DEFINITION Mus musculus 17 days embryo head cDNA, RIKEN full-length enriched library, clone:330002D11 product:EWing sarcoma homolog, full insert sequence.  
 ACCESSION AK014366  
 VERSION AK014366.1 GI:12852167  
 KEYWORDS HTC; CAP trapper.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 REFERENCE  
 AUTHORS Carninci, P. and Hayashizaki, Y.  
 TITLE High-efficiency full-length cDNA cloning  
 JOURNAL Meth. Enzymol. 303, 19-44 (1999)  
 MEDLINE 99279253  
 PUBMED 10349636  
 REFERENCE  
 AUTHORS Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Komu, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
 TITLE Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes  
 JOURNAL Genome Res. 10 (10), 1617-1630 (2000)  
 MEDLINE 20493374  
 PUBMED 11042159  
 REFERENCE  
 AUTHORS Shibata, K., Itoh, M., Aizawa, K., Nagaoaka, S., Sasaki, N., Carninci, P., Komu, H., Akiyama, J., Nishi, K., Kikunai, T., Teshiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watanabe, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J.,

TITLE	Ozekaki,Y., Muramatsu,M., Inoue,Y., Kita,A. and Hayashizaki,Y. RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer
JOURNAL	Genome Res. 10 (11), 1757-1771 (2000)
MEDLINE	PUBMED 20530913 11076861
AUTHORS	4 The RIKEN Genome Exploration Research Group Phase II Team and the PANTOM Consortium. Functional annotation of a full-length mouse cDNA collection Nature 409, 685-690 (2001) 5
TITLE	The PANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.
JOURNAL	Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
REFERENCE	Nature 420, 563-573 (2002) 6 (bases 1 to 1379)
AUTHORS	Adachi,J., Aizawa,K., Akahira,S., Altamura,T., Arai,A., Aono,H., Arakawa,T., Bono,H., Carrincci,P., Fukuda,S., Fukunishi,Y., Furuno,M., Hanagaki,T., Hara,A., Hayatsu,N., Hizamoto,K., Hirano,T., Horii,F., Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kakuoka,T., Kato,H., Kawai,Y., Kojima,Y., Komoto,H., Kouda,M., Koyas,S., Kurihara,C., Matuyama,T., Miyazaki,A., Nishii,K., Nomura,K., Numasaki,R., Ohno,M., Ozekaki,Y., Okido,T., Owa,C., Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H., Saeki,D., Shibata,K., Shidata,Y., Shingawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Tanaka,T., Tejima,Y., Toya,T., Yamamura,T., Yasunishi,A., Yoshida,K., Yoshiho,M., Muramatsu,M. and Hayashizaki,Y.
TITLE	Direct Submission
JOURNAL	Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration and Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.jp, URL:http://genome.gsc.riken.jp/, Tel:81-45-503-9222, Fax:81-45-503-9216)
COMMENT	Please visit our web site ( <a href="http://genome.gsc.riken.jp/">http://genome.gsc.riken.jp/</a> ) for further details. cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. First strand cDNA was primed with a primer [5' GAGAGAGACAGATCCAGAGAGCTTTTCTTTTTTTTTTNN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGATTTCGATTAATTAATTAATTAATCCCCCCC 3']. cDNA was cleaved with XhoI and SacI. Cloning sites, 5' end: XhoI, 3' end: SacI. Host: SOLR.
FEATURES	location/Qualifiers 1..1379 /organism="Mus musculus" /mol_type="mRNA" /strain="C57BL/6J" /db_xref="PANTOM.DB:3300002D11" /db_xref="taxon:10090" /cclone="3300002D11" /tissue_type="head" /cclone_lib="RIKEN full-length enriched mouse cDNA library" /dev_stage="17 days embryo" <1..1004 /note="unnamed protein product; Ewing sarcoma homolog (MDM); NCBI:99960, GB NM_007968, evidence: BLASTN, 99%, match=2172) putative" /codon_start=3. /protein_id="BAB29301.1" /db_xref="NCI:12652168" /translatation="LGSAGEGGCGNRKRGPMDSGLPLDPDESDNSAIVYOG LNDVTLDDLDLDFKKCGGVGNVRRTQPMHVIHLDETGPXKDAIVSYEDPEPTAA

polya_signal		polya_site		ORIGIN	
Alignment Scores:					
Pred. No.:	4.95e-105	Length:	1379		
Score:	1874.00	Matches:	328		
Percent Similarity:	99.70%	Conserved:	4		
Best Local Similarity:	98.50%	Mismatches:	1		
Query Match:	51.58%	Indels:	0		
DB:	3	Gaps:	0		
US-10-791-017A-2 (1-656) x AK014366 (1-1379)					
QY	324	MetGlySerAlaGlyIuaRgLygLyPheAsnLyeProGlygLyPromeTaBpGluGly	343		
DB	3	CTGGGCAAGCCTGGAGAGCCAGGTGGCTTCATATAGCCTGGTGGACCCATGAGTAAGAGGA	62		
QY	344	ProAspLeuAspLeuGlyProProValAspProAspLuuAspSerAspAsnSerAlaIle	363		
DB	63	CCAAATCTTGATCTAGGCGCTTCATATAGATCCCATATAGACTCTTGACCAACATGCAATT	122		
QY	364	TyrValGlnGlyLeuAsnAspSerValThrLeuAspAspLeuAlaAspPhePheGln	383		
DB	123	TATGTGCAAGATTAAATGACAATGTACTCTGATATATCTGGACAGCTTTTAAGCAG	182		
QY	384	CysGlyValValLysMetAsnLysArgThrGlyGlnProMetIleHisIleTyrLeuAsp	403		
DB	183	TGTGGGGTGTCAAGATGACACAGAGAACATGACAAACCATGATCCATATCTACCTGGAT	242		
QY	404	LysGluThrGlyLysProLysGlyAspAlaThrValSerTyrGluAspProProThrAla	423		
DB	243	AAGAGACAGGAAAGCCTTAAAGGGGAGCCACAATGTCTTATGAAAGATCCACCAACTGCA	302		
QY	424	LysAlaIaValaGluTTPheAspGlyLysAspPheGlnGlySerLysLeuLysValSer	443		
DB	303	AAGGCTGCGGTGCAATGGTTTGATGGAAAGATTTTCAAGAGCAACTTAAAGTGTCT	362		
QY	444	LeuAlaArgLysLysProPromeLAsnSerMetArgGlyGlyLeuProProArgGluGly	463		
DB	363	CTTCCCGGAAAGAGCTCCATATGACACATGGGGAGGACATGCCACTCGTAGAGGC	422		
QY	464	ArgGlyMetProProLysAspArgGlyLysProGlyLysProGlyLysProGlyLysPro	483		
DB	423	AGGGGAGTGCACACACACTTCGTGGAGATCTGTGTGGCCACAGGAGGCGCTCGAGAGCC	482		
QY	484	MetClyArgMetGlyLysArgGlyLysAspArgGlyLysPheProAspArgGlyProArg	503		
DB	483	ATGGGTGCAATGGGTGGCCGTGGAGAGACAGAGGGGGCTTCCCTCAAGAGGGCCCCGA	542		
QY	504	GlySerArgLysAsnProSerGlyGlyLysAsnValGlnHisArgAlaGlyAspTyrGln	523		
DB	543	GGCTCCAGAGAAACCCCTCTGAGAGAGAAATTCACGACACCGAGCTGAGACTGGCAG	602		
QY	524	CysProAsnProGlyLysGlyAsnGlnAsnPheAlaTyrArgThrGlyLysAsnGlnCys	543		
DB	603	TGTCCCAATCCGGGCTGTGGAAACCAAGAACTTCGCTTGAGAAACAGAAATCAACCAATGT	662		
QY	544	LysAlaProLysProGlyLysPheLeuProProProPheProProProProGlyLysAspArg	563		
DB	663	AAGGCCCCCTAAGCCCGAGGCGCTTCCTCCGCAACCTTTTCACTCCGGGTGGATCGT	722		
QY	564	GlyArgGlyLysProGlyLysMetArgGlyLysArgGlyLysLeuMetAspArgGlyLys	583		
DB	723	GGAAGAGGTGGCCCTGTGTGCATCCAGAGAGAAAGAGAGACTCATGACACGTGTGTGT	782		
QY	584	ProGlyLysMetPheArgGlyLysArgGlyLysAspArgGlyLysPheArgGlyLysArg	603		

Db 783 CCTGGAGGAAATGTTCAAGAGGTGGCAGAGGTGAGACAGAGAGGCTTCCGAGGTGGCCGT 842  
QY 604 GLYMeCAspArgGlyGlyPheGlyGlyGlyArgArgGlyGlyProGlyGlyProProGly 623  
Db 843 GGAATGACCGAGGTGCTTTGTGTGAGAGACGAGGTGTCTCTGGGGGCTCTCTGGA 902  
QY 624 ProLeuMeGluGluMeGlyGlyArgArgGlyGlyArgGlyGlyProGlyGlyProGly 643  
Db 903 CCTTTAATGAAACGATGGAT 962  
QY 644 LysGlyGluHisArgGlnGluArgArgAspArgProTyr 656  
Db 963 AAAGCGAGCACCGTCAGAAACGAGAGACCGGCCCTTAC 1001

Search completed: February 21, 2005, 04:11:40  
Job time : 4977 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 18, 2005, 15:13:01 ; Search time 165 Seconds

(without alignments)  
792.273 Million cell updates/sec

Title: US-10-791-017a-2\_COPY\_319\_656

Perfect score: 1922  
Sequence: 1 GGRGCMGSGERGFNKP...GGRGMDKGRHQRDRRY 338

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

A\_Geneseq\_16Dec04:\*  
1: geneseq19808:\*  
2: geneseq19908:\*  
3: geneseq20008:\*  
4: geneseq20018:\*  
5: geneseq20028:\*  
6: geneseq20038:\*  
7: geneseq20038:\*  
8: geneseq20048:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1922	100.0	600	ADA55145	ADA55145 Human pro
2	1922	100.0	656	ADA516278	ADA516278 Human Bv1
3	1922	100.0	656	ABM82330	ABM82330 Tumour-as
4	1922	100.0	665	ABG06460	ABG06460 Novel hum
5	1913	99.5	656	AAR44555	AAR44555 Human EMS
6	1909	99.3	604	ADR09402	ADR09402 Human pro
7	1900	98.9	656	AAW33813	AAW33813 Human EMS
8	1888.5	98.3	655	ABBS7126	ABBS7126 Human EMS
9	1810	94.2	583	ADP56334	ADP56334 Mouse lbc
10	1415	73.6	361	ABM80197	ABM80197 Tumour-as
11	878	45.7	525	ABG95084	ABG95084 Human tra
12	878	45.7	525	ADRI4649	ADRI4649 Human NF-
13	878	45.7	526	AAW78355	AAW78355 Human tra
14	878	45.7	526	ABG95081	ABG95081 Human tra
15	878	45.7	526	AD126117	AD126117 Human tra
16	878	45.7	526	ABM81732	ABM81732 Tumour-as
17	878	45.7	526	ADSS8302	ADSS8302 Human pro
18	878	45.7	569	AAW79339	AAW79339 Human pro
19	878	45.7	569	AAW79340	AAW79340 Human pro
20	873.5	45.4	518	AD126113	AD126113 Human pro
21	870	45.3	536	AAW78356	AAW78356 Human pro
22	799.5	41.6	260	ADJ68310	ADJ68310 Human hea
23	750	39.0	589	AD126207	AD126207 Human pro
24	750	39.0	592	ABO53050	ABO53050 Human put
25	750	39.0	592	AD126209	AD126209 Human pro

26	623	32.4	156	3	ABA44066	ABA44066 Human can
27	616.5	32.1	399	4	ABB60010	ABB60010 Drosophil
28	475	24.7	83	4	AAW21854	AAW21854 Peptide #
29	475	24.7	83	4	ABM44223	ABM44223 Peptide #
30	475	24.7	83	4	AAW38178	AAW38178 Peptide #
31	475	24.7	83	4	ABM27099	ABM27099 Protein #
32	475	24.7	83	4	AAW7960	AAW7960 Human bon
33	475	24.7	83	4	AAW55255	AAW55255 Human bra
34	475	24.7	83	4	ABG59596	ABG59596 Human liv
35	475	24.7	83	5	ABG46973	ABG46973 Human pep
36	419	21.8	266	7	ADL22735	ADL22735 Human dis
37	407.5	21.2	280	8	AD126115	AD126115 Human pro
38	303	15.8	1078	2	AAW1704	AAW1704 Collagen
39	303	15.8	1078	3	AAW96125	AAW96125 Collagen
40	303	15.8	1078	5	AAE16478	AAE16478 Human col
41	303	15.8	1078	5	ABM80736	ABM80736 Collagen
42	303	15.8	1078	5	ABM80628	ABM80628 Amino aci
43	303	15.8	1078	7	ADFI3078	ADFI3078 Human col
44	303	15.8	1466	4	AAE02534	AAE02534 Bovine al
45	303	15.8	1466	4	AAE02533	AAE02533 Bovine al

#### ALIGNMENTS

RESULT 1  
ID ADA55145  
ADA55145 standard; protein; 600 AA.  
XX  
AC ADA55145;  
XX  
DT 20-NOV-2003 (first entry)  
XX  
DE Human protein, SEQ ID 2713.  
XX  
DE  
XX  
XX Cytotoxic; Anti-inflammatory; Osteopathic; Neuroprotective; Nootropic;  
XX Gene Therapy; human; secretory protein; membrane proteins; cancer;  
XX Inflammatory disease; osteoporosis; neurological disease.  
XX  
XX Homo sapiens.  
XX  
XX EP1293569-A2.  
XX  
XX 19-MAR-2003.  
XX  
XX 21-MAR-2002; 2002EP-00006586.  
XX  
XX 14-SEP-2001; 2001JP-00328381.  
XX  
XX 24-JAN-2002; 2002US-0350435P.  
XX  
XX (HELI-) HELIX RES INST.  
XX (REAS-) RES ASSOC BIOTECHNOLOGY.  
XX  
XX Isogai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S;  
XX Yamamoto J, Isono Y, Hio Y, Otsuka K, Negai K, Irie R, Tamechika I;  
XX Seki N, Yoshikawa T, Otsuka M, Nagahara K, Masuno Y;  
XX WPI; 2003-395539/38.  
XX DR N-PSDB; ADA53506.  
XX  
XX New polynucleotides encoding full-length polypeptides, e.g. secretory  
XX and/or membrane proteins, useful for developing medicines for diseases in  
XX which the gene is involved, or as target molecules for gene therapy.  
XX  
XX Claim 14; SEQ ID NO 2713; 205pp; English.  
XX  
XX The present invention relates to novel human secretory or membrane  
XX proteins (ADA54072-ADA55710) and their coding sequences (ADA52433-  
XX ADA54071). The coding sequences are useful in the gene therapy of  
XX diseases caused by abnormalities of the proteins, e.g. cancer,  
XX inflammatory diseases, osteoporosis or neurological disease.  
XX  
XX Sequence 600 AA:

Query Match 100.0%; Score 1922; DB 6; Length 600;  
Best Local Similarity 100.0%; Pred. No. 5,4e-148;  
Matches 338; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGRGMSAGRGSGFNNKPGPMDEGPDLDLGPVDPDESDNSAIYVQGLNSVTLDDLA 60  
DB 263 GGRGMSAGRGSGFNNKPGPMDEGPDLDLGPVDPDESDNSAIYVQGLNSVTLDDLA 322  
QY 61 DFFKCGVVKNNKRTGQPMIHIYLDKETGPKGDAIVSYEDPPTAKAAVWFPGKDFGGS 120  
DB 323 DFFKCGVVKNNKRTGQPMIHIYLDKETGPKGDAIVSYEDPPTAKAAVWFPGKDFGGS 382  
QY 121 KLVSLARKKPPMNSNRGGLPPREGKMPPLRGPGGPGGPGMGRMGSGDGGGPP 180  
DB 383 KLVSLARKKPPMNSNRGGLPPREGKMPPLRGPGGPGGPGMGRMGSGDGGGPP 442  
QY 181 PRGPRGSRGNPSGGGVVQHRAGDMQCPNPGCGNQNFAMRTGKQCAKPEEGFLPPPPPP 240  
DB 443 PRGPRGSRGNPSGGGVVQHRAGDMQCPNPGCGNQNFAMRTGKQCAKPEEGFLPPPPPP 502  
QY 241 PGDRGRGGPGGMRGGRGGLMDRGPGGMRGGRGGDRGGFRGGRGMDRGFGGGRGGP 300  
DB 503 PGDRGRGGPGGMRGGRGGLMDRGPGGMRGGRGGDRGGFRGGRGMDRGFGGGRGGP 562  
QY 301 GGPPLMEQMGRRGGRGGPGKMDKGEHROBRDRPY 338  
DB 563 GGPPLMEQMGRRGGRGGPGKMDKGEHROBRDRPY 600

RESULT 2  
ADSI6278  
ID ADSI6278 standard; protein; 656 AA.  
AC ADSI6278;  
XX  
DT 18-NOV-2004 (first entry)  
DE Human Ewing sarcoma protein SegID 2.  
XX  
KW human; Ewing sarcoma; EWS; prostatic cancer; alopecia; acne;  
KW hypogonadism; androgen-resistance syndrome; testicular feminization.  
XX  
OS Homo sapiens.  
XX  
PN BPI455190-A1.  
PD 08-SEP-2004.  
XX  
PE 16-FEB-2004; 2004EP-00003422.  
XX  
PR 04-MAR-2003; 2003DE-01009280.  
XX  
PR 25-APR-2003; 2003US-0465692P.  
XX  
PA (SCHD ) SCHERING AG.  
XX  
PI Obendorf M, Wolf S;  
XX  
DR WPI; 2004-627861/61.  
DR N-PSDB; ADSI6277.  
XX  
XX  
PT Determining the hormonal effects of substances, used to identify  
PT pharmaceuticals, e.g. for treatment of androgen receptor dysfunction,  
PT from modulating interaction between nuclear receptors and Ewing sarcoma  
PT protein.  
XX  
PS Disclosure; SEQ ID NO 2; 30pp; German.  
XX  
XX This invention relates to a novel modulators that alter the interaction  
XX between the Ewing sarcoma protein (EWS) and its nuclear receptor, as well  
XX as the screening method thereof. Specifically, it refers to determining  
XX and identifying a hormonal effect brought about by test compounds that  
XX modulate either the binding of EWS to the nuclear receptor or the ligand-

CC induced activity of this receptor. The present invention describes the  
CC nuclear receptors as including oestrogen, progesterone, thyroid hormone,  
CC vitamin D, and retinoic acid receptors, most preferably they are androgen  
CC receptors. Accordingly, these modulators may be used in the development  
CC of pharmaceutical compositions that can diagnose and be used to treat  
CC diseases associated with receptor dysfunction such as prostatic cancer,  
CC alopecia, acne, hypogonadism and androgen-resistance syndrome e.g.  
CC testicular feminization. This method provides reliable, sensitive,  
CC simple, inexpensive and rapid assessment of the hormonal effects of these  
CC test compounds. This polypeptide sequence is the human Ewing sarcoma  
CC protein of the invention.  
XX  
SQ Sequence 656 AA;

Query Match 100.0%; Score 1922; DB 8; Length 656;  
Best Local Similarity 100.0%; Pred. No. 6e-148;  
Matches 338; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGRGMSAGRGSGFNNKPGPMDEGPDLDLGPVDPDESDNSAIYVQGLNSVTLDDLA 60  
DB 319 GGRGMSAGRGSGFNNKPGPMDEGPDLDLGPVDPDESDNSAIYVQGLNSVTLDDLA 378  
QY 61 DFFKCGVVKNNKRTGQPMIHIYLDKETGPKGDAIVSYEDPPTAKAAVWFPGKDFGGS 120  
DB 379 DFFKCGVVKNNKRTGQPMIHIYLDKETGPKGDAIVSYEDPPTAKAAVWFPGKDFGGS 438  
QY 121 KLVSLARKKPPMNSNRGGLPPREGKMPPLRGPGGPGGPGMGRMGSGDGGGPP 180  
DB 439 KLVSLARKKPPMNSNRGGLPPREGKMPPLRGPGGPGGPGMGRMGSGDGGGPP 498  
QY 181 PRGPRGSRGNPSGGGVVQHRAGDMQCPNPGCGNQNFAMRTGKQCAKPEEGFLPPPPPP 240  
DB 499 PRGPRGSRGNPSGGGVVQHRAGDMQCPNPGCGNQNFAMRTGKQCAKPEEGFLPPPPPP 558  
QY 241 PGDRGRGGPGGMRGGRGGLMDRGPGGMRGGRGGDRGGFRGGRGMDRGFGGGRGGP 300  
DB 559 PGDRGRGGPGGMRGGRGGLMDRGPGGMRGGRGGDRGGFRGGRGMDRGFGGGRGGP 618  
QY 301 GGPPLMEQMGRRGGRGGPGKMDKGEHROBRDRPY 338  
DB 619 GGPPLMEQMGRRGGRGGPGKMDKGEHROBRDRPY 656

RESULT 3  
ABM82330  
ID ABM82330 standard; protein; 656 AA.  
AC ABM82330;  
XX  
DT 18-NOV-2004 (first entry)  
DE Tumour-associated antigenic target (TAT) polypeptide PRO58232, SEQ:5990.  
XX  
XX Tumour-associated antigenic target; TAT; human; overexpression; cancer;  
XX tumour; diagnosis; cell proliferative disorder; breast cancer;  
XX colorectal cancer; lung cancer; ovarian cancer; liver cancer;  
XX central nervous system cancer; bladder cancer; pancreatic cancer;  
XX cervical cancer; melanoma; leukaemia; hybridisation probe;  
XX chromosome identification; leukaemia; hybridisation probe;  
XX gene therapy; cytostatic.  
XX  
OS Homo sapiens.  
XX  
XX  
XX  
XX WO2004030615-A2.  
XX  
XX 15-APR-2004.  
XX  
XX 29-SEP-2003; 2003WO-US028547.  
XX  
XX 02-OCT-2002; 2002US-0414971P.  
XX  
XX (GETH ) GENENTECH INC.  
XX  
XX

PI Wu TD, Zhang Z, Zhou Y;  
XX WPI, 2004-347921/32.  
DR N-PSDB; ACN40903.  
XX  
XX New tumor-associated antigenic target polypeptides and nucleic acids,  
PT useful in preparing a medicament for treating or detecting a  
PT proliferative disorder, e.g. breast, lung, colorectal, ovarian or  
PT prostate cancer or tumor.  
XX  
XX Claim 12; SEQ ID NO 5990; 7273bp; English.  
XX  
XX The invention relates to human tumour-associated antigenic target (TAT)  
CC polypeptides, and their related nucleic acids. The TAT polypeptides are  
CC overexpressed in cancer tissues compared to normal tissues, and may thus  
CC serve as effective targets for the diagnosis and treatment of cancer in  
CC mammals. The invention also relates to nucleic acid and polypeptide  
CC sequences at least 80% identical to the TAT nucleic acids and  
CC polypeptides; expression vectors and host cells comprising a TAT nucleic  
CC acid, an antibody specific for a TAT polypeptide; a peptide or organic  
CC molecule which binds to a TAT polypeptide; fusion proteins comprising a  
CC TAT polypeptide; and methods and compositions for the treatment or  
CC diagnosis of cancer in mammals. TAT polypeptides, nucleic acids,  
CC antibodies, antagonists, binding molecules and compositions are useful  
CC for diagnosing or treating a cell proliferative disorder associated with  
CC increased TAT expression, particularly cancers such as breast cancer,  
CC colorectal cancer, lung cancer, ovarian cancer, liver cancer, bladder  
CC cancer, pancreatic cancer, cervical cancer, cancers of the central  
CC nervous system, melanoma and leukaemia. TAT nucleic acids may further be  
CC used as hybridisation probes, in chromosome and gene mapping, in  
CC chromosome identification and in gene therapy. The present sequence  
CC represents a TAT polypeptide of the invention  
XX  
XX  
SQ Sequence 656 AA;  
Query Match 100.0%; Score 1922; DB 8; Length 656;  
Best Local Similarity 100.0%; Pred. No. 6e-148; Indels 0; Gaps 0;  
Matches 338; Conservative 0; Mismatches 0;  
QY 1 GGRGSGSAGRGFGFKPGGPMDEGPDLDGPPVDEDESDNSAIYVGLNDSVTLLDLA 60  
DB 319 GGRGSGSAGRGFGFKPGGPMDEGPDLDGPPVDEDESDNSAIYVGLNDSVTLLDLA 378  
QY 61 DFFKQCGVVYKMKRTGQPMIHIYLDKGTGPKGDATVSYEDPPTAKAAVEMFPGKDFQGS 120  
DB 379 DFFKQCGVVYKMKRTGQPMIHIYLDKGTGPKGDATVSYEDPPTAKAAVEMFPGKDFQGS 428  
QY 121 KLVSLARKKKPPMNSMRGGLPREGRGMPPLLRGSGGPGGPGMGRMGSGDGGFP 180  
DB 439 KLVSLARKKKPPMNSMRGGLPREGRGMPPLLRGSGGPGGPGMGRMGSGDGGFP 498  
QY 181 PRGPRGSRGNPGSGGVNORAGDMQCPNPGCGNFMFAMTECNQCAKPEBGLPPFPFP 240  
DB 499 PRGPRGSRGNPGSGGVNORAGDMQCPNPGCGNFMFAMTECNQCAKPEBGLPPFPFP 558  
QY 241 PGSDRGRGPGGMRGGRGGLMDRGPGQMFRRGGRGGDGGFRGGRMDRGFGGSGRGGP 300  
DB 559 PGSDRGRGPGGMRGGRGGLMDRGPGQMFRRGGRGGDGGFRGGRMDRGFGGSGRGGP 618  
QY 301 GGPPLMEOMGGRGGRGGGPGKMDKGEHRQERRDRPY 338  
DB 619 GGPPLMEOMGGRGGRGGGPGKMDKGEHRQERRDRPY 656

XX  
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder.  
XX  
XX Homo sapiens.  
OS  
XX WO200175067-A2.  
PN  
XX 11-OCT-2001.  
PD  
XX  
XX 30-MAR-2001; 2001WO-US008631.  
PF  
XX  
XX 31-MAR-2000; 2000US-00540217.  
PR 23-AUG-2000; 2000US-00649167.  
XX  
XX (HYSE-) HYSEQ INC.  
PA  
XX Dmanac RT, Liu C, Tang YT;  
PI  
XX WPI, 2001-639362/73.  
DR N-PSDB; AAS70647.  
XX  
XX New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity.  
XX  
XX Claim 20; SEQ ID NO 36819; 103bp; English.  
XX  
XX The invention relates to isolated polynucleotide (I) and polypeptide (II)  
CC sequences. (I) is useful as hybridisation probes, polymerase chain  
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,  
CC and in recombinant production of (II). The polynucleotides are also used  
CC in diagnostics as expressed sequence tags for identifying expressed  
CC genes. (I) is useful in gene therapy techniques to restore normal  
CC activity of (II) or to treat disease states involving (II). (II) is  
CC useful for generating antibodies against it, detecting or quantitating a  
CC polypeptide in tissue, as molecular weight markers and as a food  
CC supplement. (II) and its binding partners are useful in medical imaging  
CC of sites expressing (II). (I) and (II) are useful for treating disorders  
CC involving aberrant protein expression or biological activity. The  
CC polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. ABG0010-ABG30377 represent novel human diagnostic  
CC amino acid sequences of the invention. Note: The sequence data for this  
CC patent did not appear in the printed specification, but was obtained in  
CC electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pat\_sequences  
XX  
XX  
SQ Sequence 665 AA;  
Query Match 100.0%; Score 1922; DB 4; Length 665;  
Best Local Similarity 100.0%; Pred. No. 6.1e-148; Indels 0; Gaps 0;  
Matches 338; Conservative 0; Mismatches 0;  
QY 1 GGRGSGSAGRGFGFKPGGPMDEGPDLDGPPVDEDESDNSAIYVGLNDSVTLLDLA 60  
DB 328 GGRGSGSAGRGFGFKPGGPMDEGPDLDGPPVDEDESDNSAIYVGLNDSVTLLDLA 387  
QY 61 DFFKQCGVVYKMKRTGQPMIHIYLDKGTGPKGDATVSYEDPPTAKAAVEMFPGKDFQGS 120  
DB 388 DFFKQCGVVYKMKRTGQPMIHIYLDKGTGPKGDATVSYEDPPTAKAAVEMFPGKDFQGS 447  
QY 121 KLVSLARKKKPPMNSMRGGLPREGRGMPPLLRGSGGPGGPGMGRMGSGDGGFP 180  
DB 448 KLVSLARKKKPPMNSMRGGLPREGRGMPPLLRGSGGPGGPGMGRMGSGDGGFP 507  
QY 181 PRGPRGSRGNPGSGGVNORAGDMQCPNPGCGNFMFAMTECNQCAKPEBGLPPFPFP 240  
DB 508 PRGPRGSRGNPGSGGVNORAGDMQCPNPGCGNFMFAMTECNQCAKPEBGLPPFPFP 567

Oy		241	PGCDRGRGCGCNRGCGGLMDRGCPGCFRGCRCGDRGCRGCMRDGCGFGGRRGCGP	300
Dd		568	PGCDRGRGCGCNRGCGGLMDRGCPGCFRGCRCGDRGCRGCMRDGCGFGGRRGCGP	627
Oy		301	GGPFGPLMEQMGRRGCGGPGKMDXKEHROERRDRPY	338
Dd		628	GGPFGPLMEQMGRRGCGGPGKMDXKEHROERRDRPY	665
	RESULT 5			
ID	AAR44555			
XX	AAR44555 standard; protein; 656 AA.			
XX	AAR44555;			
XX	AC			
XX	DT 25-MAR-2003 (revised)			
DT	26-MAY-1994 (first entry)			
XX				
DE	Human EMS protein deduced from foetal brain cDNA clone BFLA5.			
XX				
KW	chromosomal translocation; chimeric; chimaeric; Ewing sarcoma; Ews gene;			
KM	malignant melanoma; human chromosome 11;			
KV	primitive peripheral neuroectodermal tumour; human chromosome 22;			
KW	hum-flt-1.			
XX				
OS	Homo sapiens.			
XX				
XX	Key Location/Qualifiers			
FH	Misc-difference 569			
FT	/note= "corresponds to GGT codon"			
XX				
PN	WO9323549-A2.			
XX				
PD	25-NOV-1993.			
XX				
PF	19-MAY-1993; 93MO-FR000494.			
XX				
PR	20-MAY-1992; 92FR-00006123.			
XX				
PA	(CNRS ) CNRS CENT NAT RECH SCI.			
PI	Aurias A, Delattre O, Desmaze C, Melot T, Peter M, Plougaestel B;			
PT	Thomas G, Zucman J;			
XX				
DR	WPI; 1993-386580/48.			
XX				
DR	N-PSDB; AAQ50643.			
XX				
PT	New nucleic acid of EWS gene and its hybrid(s) - contg. gene sequence			
XX	involved in chromosomal trans-location, also derived mRNA, probes, fusion			
XX	proteins etc., for diagnosis and treatment of Ewing sarcoma and melanoma.			
XX				
PS	Disclosure; Flg 6; 123pp; French.			
CC	The probes 22RR3 and 22RR12 were used to screen a human foetal brain cDNA			
CC	library (Stratagene cat.# 936206). The clone bFLA5 was identified and			
CC	sequenced; it represents the entire coding region and 3'-UTR of the Ews			
CC	gene. (Updated on 25-MAR-2003 to correct PN field.)			
XX				
SQ	Sequence 656 AA:			
	Query Match 99.5%; Score 1913; DB 2; Length 656;			
	Bee Local Similarity 99.7%; Pred. No.3.2e-147;			
	Matches 337; Conservative 0; Mismatches 1; Indels 0; Gaps 0			
Oy		1	GGRGCMGSAGEGRCGFPNKGPPMDGEPDLDPGPPVPDDEDSDNSATYVGILNDSTLDDLA	60
Dd		319	GGRGCMGSAGEGRCGFPNKGPPMDGEPDLDPGPPVPDDEDSDNSATYVGILNDSTLDDLA	378
Oy		61	DFFRCQGVVMNKRTSQPMIHITLDKETGPKYGDA TVSYEDPPTAKAAVEWFDCKDFQS	120
Dd		379	DFFRCQGVVMNKRTSQPMIHITLDKETGPKYGDA TVSYEDPPTAKAAVEWFDCKDFQS	438
Oy		121	KLVSLARKRPPEMNWRGCLPPEEGRGMPPLRGCBGCGCGGPMGMGSGRGDRGGFP	180

Qy	439	KLKVSLARKKKPPMNSMRGLPPREGRGMPPLRLRGPGGPGGPGMGRGRGDDGSGP	498
Db	181	PRGPRGSRNPPGGGNNQHRAGMDQCPRPGCGNQNFMARTECNQCKRPREGFLPPRPFP	240
Db	499	PRGPRSRNPNPGGNGVHRAGDWCQCPRGCGNQNFMARTECNQCKRPREGFLPPRPFP	558
Qy	241	PGDGRGRGPGGMRGGRGGLMDRGPGGMPFRGGRGGRGPRGGRGMDRGFGGGRG	300
Db	559	PGDGRGRGPGVGMRRGRGGLMDRGPGGMPFRGGRGGRGPRGGRGMDRGFGGGRG	618
Qy	301	GGPPGLMEQMGRRGRGGRGPGKMDKGRHQRERRDRY	338
Db	619	GGPPGLMEQMGRRGRGGRGPGKMDKGRHQRERRDRY	656
RESULT 6			
ADRO9402			
ADRO9402		standard; protein; 604 AA.	
AC	ADRO9402;		
XX			
DT	04-NOV-2004	(first entry)	
XX			
DE		Human protein useful for treating neurological disease Seq 2908.	
XX			
KW	human; oligo-capping method; diagnostic marker; gene therapy;		
KW	osteoporosis; neurological disease; Alzheimer's disease;		
KW	Parkinson's disease; dementia; short memory; cancer;		
KW	sense or motor function; emotional reaction; fear response; panic;		
KW	osteopathic; neuroprotective; nootropic; antiparkinsonian; cyostatic;		
KW	tranquilliser.		
XX			
OS	Homo sapiens.		
PN	EP1447413-A2.		
XX			
PD	18-AUG-2004.		
PF	12-FEB-2004; 2004EP-00003145.		
XX			
PR	14-FEB-2003; 2003JP-00102207.		
XX	09-MAY-2003; 2003JP-00131452.		
PA	(REAS-) RES ASSOC BIOTECHNOLOGY.		
PI	Isogai T, Yamamoto J, Nishikawa T, Isono Y, Sugiyama T, Otsuki T;		
PI	Wakamatsu A, Ishii S, Nagai K, Irie R;		
XX	WPI; 2004-583265/57.		
DR	N-PSDB; ADR07446.		
PT	New 1995 cDNA, useful for treating osteoporosis, neurological diseases,		
PT	Alzheimer's diseases, Parkinson's diseases, dementia and various cancers.		
PS	Claim 1, SEQ ID NO 2908; 2686bp; English.		
XX			
CC	This invention relates to novel, isolated full length human cDNA		
CC	molecules and the encoded proteins thereof. Specifically, it refers to		
CC	cDNA clones obtained by an oligo-capping method, where none of these		
CC	clones are identical to any known human mRNAs. The present invention		
CC	describes an immunoassay to identify agonists and antagonists, as well as		
CC	antibodies, antisense molecules and siRNAs that can all be used to bind		
CC	to and modulate expression of the cDNA molecules. As such, these		
CC	molecules are useful for diagnostic markers or therapeutic targets for		
CC	the various diseases or morbid states. In particular, they are useful in		
CC	gene therapy for treating osteoporosis, neurological disease, Alzheimer's		
CC	disease, Parkinson's disease, dementia, short memory and various cancers,		
CC	as well as for maintaining equilibrium of sense or motor function, and		
CC	for treating emotional reaction, fear response and panic. Accordingly,		
CC	they exhibit osteopathic, neuroprotective, nootropic, antiparkinsonian,		
CC	cyostatic and tranquilliser activities. This polypeptide is a protein		
CC	encoded by a full length human cDNA sequence of the invention. NOTE: This		

CC sequence is not given in the sequence listing of the specification but  
CC can be obtained on CD-ROM from the European Patent Office, Vienna Sub-  
CC office.

XX Sequence 604 AA;

Query Match 99.3%; Score 1909; DB 8; Length 604;  
Best Local Similarity 99.7%; Pred. No. 6.2e-147;  
Matches 337; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GGRGMSAGRGSGFNRKPGPMDEGPDLDGPPVDDESDNSAIYVQGLNDSVTLDDLA 60  
DB 267 GGRGMSAGRGSGFNRKPGPMDEGPDLDGPPVDDESDNSAIYVQGLNDSVTLDDLA 326  
QY 61 DFFKQCGVVMKRRGTQPMIHLYLDKETGPKGDATVSYEDPPTAKAAVEMFDGKDFGS 120  
DB 327 DFFKQCGVVMKRRGTQPMIHLYLDKETGPKGDATVSYEDPPTAKAAVEMFDGKDFGS 386  
QY 121 KLVSLARKKPPNNSMRGGLPREGGMPPLRGPGGPGGPMGRMGGRGDRGDF 180  
DB 387 KLVSLARKKPPNNSMRGGLPREGGMPPLRGPGGPGGPMGRMGGRGDRGDF 446  
QY 181 PRGPRGSRGNPSGGGVVQHRAGDMQCPNPGCGNPNFAMRTGCKQKAPKEGFLPPFP 240  
DB 447 PRGPRGSRGNPSGGGVVQHRAGDMQCPNPGCGNPNFAMRTGCKQKAPKEGFLPPFP 506  
QY 241 PGDRGRGPGGMRGGRGGLMDRGPGGMFRGGRGDRGFRGGRGMDRGFGGGRG 300  
DB 507 PGDRGRGPGGMRGGRGGLMDRGPGGMFRGGRGDRGFRGGRGMDRGFGGGRG 566  
QY 301 GGPPLMEQMGRRGGRGPGKMDKGEHRQERRDRPY 338  
DB 567 GGPPLMEQMGRRGGRGPGKMDKGEHRQERRDRPY 604

RESULT 7

AAW33813 standard; protein; 656 AA.

XX ID AAW33813  
XX AC AAW33813  
XX DT 06-JUL-1998 (first entry)  
XX DE Human EMS protein.  
XX KW Tat stimulatory factor; Tat-SF1; transcriptional activity factor; HIV-1;  
XX infection; gene therapy; EMS.  
XX OS Homo sapiens.  
XX FH Key  
XX FT Region  
XX FT 30..44  
XX FT /note= "Imperfect repeat peptide showing homology to Tat-SF1"  
XX FT Region  
XX FT 209..236  
XX FT /note= "Imperfect repeat peptide showing homology to Tat-SF1"  
XX PN WO9800695-A2.  
XX PD 08-JAN-1998.  
XX PF 03-JUL-1997, 97WO-US011713.  
XX PR 03-JUL-1996; 96US-0021218P.  
XX PR 13-DEC-1996; 96US-0033152P.  
XX PA (MAST) MASSACHUSETTS INST TECHNOLOGY.  
XX PI Sharp PA, Zhou Q;  
XX DR WPI, 1998-087086/08.

PT Nucleic acid encoding Tat stimulating factor protein and related  
PT transformed cells - proteins and binding agents, used to treat human  
PT immunodeficiency virus infection.

XX Example 6; Page 47-51; 68pp; English.

CC EMS is a member of a novel family of putative transcription factors that  
CC have RNA recognition motifs and which are frequently associated with many  
CC types of sarcomas. It shows homology to a novel human transcriptional  
CC activity factor, Tat-stimulatory factor (Tat-SF1, see AAW33813), that is  
CC involved in the regulation of transcriptional elongation of HIV-1 by Tat.  
CC is essential for Tat trans-activation and is a substrate of an associated  
CC cellular kinase. Tat-SF can be used to screen for binding agents useful  
CC in the treatment of HIV infection

XX Sequence 656 AA;

Query Match 98.9%; Score 1900; DB 2; Length 656;  
Best Local Similarity 99.4%; Pred. No. 3.7e-146;  
Matches 336; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GGRGMSAGRGSGFNRKPGPMDEGPDLDGPPVDDESDNSAIYVQGLNDSVTLDDLA 60  
DB 319 GGRGMSAGRGSGFNRKPGPMDEGPDLDGPPVDDESDNSAIYVQGLNDSVTLDDLA 378  
QY 61 DFFKQCGVVMKRRGTQPMIHLYLDKETGPKGDATVSYEDPPTAKAAVEMFDGKDFGS 120  
DB 379 DFFKQCGVVMKRRGTQPMIHLYLDKETGPKGDATVSYEDPPTAKAAVEMFDGKDFGS 438  
QY 121 KLVSLARKKPPNNSMRGGLPREGGMPPLRGPGGPGGPMGRMGGRGDRGDF 180  
DB 439 KLVSLARKKPPNNSMRGGLPREGGMPPLRGPGGPGGPMGRMGGRGDRGDF 498  
QY 181 PRGPRGSRGNPSGGGVVQHRAGDMQCPNPGCGNPNFAMRTGCKQKAPKEGFLPPFP 240  
DB 499 PRGPRGSRGNPSGGGVVQHRAGDMQCPNPGCGNPNFAMRTGCKQKAPKEGFLPPFP 558  
QY 241 PGDRGRGPGGMRGGRGGLMDRGPGGMFRGGRGDRGFRGGRGMDRGFGGGRG 300  
DB 559 PGDRGRGPGGMRGGRGGLMDRGPGGMFRGGRGDRGFRGGRGMDRGFGGGRG 618  
QY 301 GGPPLMEQMGRRGGRGPGKMDKGEHRQERRDRPY 338  
DB 619 GGPPLMEQMGRRGGRGPGKMDKGEHRQERRDRPY 656

RESULT 8

ABB57126 standard; protein; 655 AA.

XX ID ABB57126  
XX AC ABB57126;  
XX DT 07-MAR-2002 (first entry)  
XX DE Mouse ischaemic condition related protein sequence SEQ ID NO:289.  
XX KW Mouse; ischaemia; compressive ischaemia; occlusive ischaemia;  
XX vasoepaastic ischaemia; ischaemic condition; ischaemic disease.  
XX OS Mus musculus.  
XX PN WO200188188-A2.  
XX PD 22-NOV-2001.  
XX PF 18-MAY-2001; 2001WO-JP004192.  
XX PR 18-MAY-2000; 2000JP-00145977.  
XX PA (UYNI-) UNIV NIHON SCHOOL JURIDICAL PERSON.  
XX PI Ishikawa K, Asai S, Takahashi Y, Nagata T, Ishii Y;





Db 446 AGMOCNPGCCGNQNFAMRTECNQCAKPEBGLPPPPPPGDRGSGCGMKGCGGL 505  
 QY 261 MDRGGPGCMFRGGRRGGRRGGMDRGFGGRRGGPGPPPLMEQMGRRGG 320  
 Db 506 MDRGGPGCMFRGGRRGGRRGGMDRGFGGRRGGPGPPPLMEQMGRRGG 565  
 QY 321 PGKMDGGEHROERRDRPY 338  
 Db 566 PGKMDGGEHROERRDRPY 583  
 RESULT 10  
 ABM80197  
 ID ABM80197 standard; protein; 361 AA.  
 AC ABM80197;  
 XX 18-NOV-2004 (first entry)  
 DT XX  
 DE Tumour-associated antigenic target (TAT) polypeptide PRO80700, SEQ:491.  
 XX Tumour-associated antigenic target; TAT; human; overexpression; cancer;  
 KM tumour; diagnosis; cell proliferative disorder; breast cancer;  
 KM colorectal cancer; lung cancer; ovarian cancer; liver cancer;  
 KM central nervous system cancer; bladder cancer; pancreatic cancer;  
 KM cervical cancer; melanoma; leukaemia; hybridisation probe;  
 KM chromosome identification; chromosome mapping; gene mapping;  
 XX gene therapy; cytostatic.  
 OS Homo sapiens.  
 XX  
 PN MO2004030615-A2.  
 PD 15-APR-2004.  
 XX  
 PF 29-SEP-2003; 2003MO-US028547.  
 XX  
 PR 02-OCT-2002; 2002US-0414971P.  
 XX  
 PA (GETH ) GENENTECH INC.  
 XX  
 PI Wu TD, Zhang Z, Zhou Y;  
 XX  
 DR WPI; 2004-347921/32.  
 XX  
 DR N-PSDB; ACN37537.  
 XX  
 PT New tumor-associated antigenic target polypeptides and nucleic acids,  
 PT useful in preparing a medicament for treating or detecting a  
 PT proliferative disorder, e.g. breast, lung, colorectal, ovarian or  
 PT prostate cancer or tumor.  
 XX  
 XX  
 PS Claim 12; SEQ ID NO 491; 7273bp; English.  
 XX  
 CC The invention relates to human tumour-associated antigenic target (TAT)  
 CC polypeptides, and their related nucleic acids. The TAT polypeptides are  
 CC overexpressed in cancer tissues compared to normal tissues, and may thus  
 CC serve as effective targets for the diagnosis and treatment of cancer in  
 CC mammals. The invention also relates to nucleic acid and polypeptide  
 CC sequences at least 80% identical to the TAT nucleic acids and  
 CC polypeptides; expression vectors and host cells comprising a TAT nucleic  
 CC acid; an antibody specific for a TAT polypeptide; a peptide or organic  
 CC molecule which binds to a TAT polypeptide; fusion proteins comprising a  
 CC TAT polypeptide; and methods and compositions for the treatment or  
 CC diagnosis of cancer in mammals. TAT polypeptides, nucleic acids,  
 CC antibodies, antagonists, binding molecules and compositions are useful  
 CC for diagnosing or treating a cell proliferative disorder associated with  
 CC increased TAT expression, particularly cancers such as breast cancer,  
 CC colorectal cancer, lung cancer, ovarian cancer, liver cancer, bladder  
 CC cancer, pancreatic cancer, cervical cancer, cancers of the central  
 CC nervous system, melanoma and leukaemia. TAT nucleic acids may further be  
 CC used as hybridisation probes, in chromosome and gene mapping, in  
 CC chromosome identification and in gene therapy. The present sequence

CC represents a TAT polypeptide of the invention  
 XX  
 XX Sequence 361 AA;  
 SQ  
 Query Match 73.6%; Score 1415; DB 8; Length 361;  
 Best Local Similarity 78.1%; Pred. No. 6.3e-107;  
 Matches 261; Conservative 14; Mismatches 41; Indels 18; Gaps 2;  
 QY 1 GGRGWSAGRGFPNKGPPMDGPDLDGPPVDPEDSDNSAIYQGLNDVTLDDL 60  
 Db 45 GCGGWSAGEQVGFNKPGGPMDEGPDLDGPPVDPEDSDNSAIYQGLNDVTLDDL 104  
 QY 61 DFFKGGVVMNRRRTGQPMHIYLDKETGKPKGDATSYDDPPAKAAVMPFGKDG 120  
 Db 105 DFFKGGVVMNRRRTGQPMHIYLDKETGKPKGDATVSCDSIPAKAAVMPFGKDG 164  
 QY 121 KLVSLARKRPNNNSRNGGLPPREGKMPPLRLRGPGGPGGPMRMGRRGGDRGFP 180  
 Db 165 KLVSLARKRPNNNSRNGGLPPREGKMPPLRLRGPGGPGGPMRMGRRGGDRGGLP 224  
 QY 181 PRGPRSGRNPSCGGVYVHRASGMQCPNPGCGNQNFAMRTECNQCAKPEBGLPPPPP 240  
 Db 225 PRGPRSGRNPSCGGVYVHRASGMQCPNPGCGNQNFAMRTECNQCAKPEBGLPPPPP 284  
 QY 241 PGDDRGRGPGRRGRRGGLMDRGPGCMFRGGRRGGMDRGFGGRRGGPGGRRGGP 300  
 Db 285 PGDDRGRGPGRRGRRGGLMDRGPGCMFRGGRRGGMDRGFGGRRGGPGGRRGGP 335  
 QY 301 GGPFPPLMEQMGRRGGRRGGPGKMDGGEHROERR 334  
 Db 336 -----LVER--DKVALGSPRTFVTNGRKRR 360  
 RESULT 11  
 ABG95084  
 ID ABG95084 standard; protein; 525 AA.  
 XX  
 AC ABG95084;  
 XX  
 DT 04-DEC-2002 (first entry)  
 XX  
 DE Human translocation (12; 16)(q13; p11) protein #5.  
 XX  
 KM Chromosome aberration; oncogenic fusion protein; cancer;  
 KM proliferative disease; cellular protein isoform; heat shock protein 90;  
 KM HSP-90; rheumatoid arthritis; cancer; haematopoietic disorder;  
 KM T cell lymphoma; B cell lymphoma; chronic myeloid leukaemia; CML;  
 KM acute myeloid leukaemia; AML; chronic myelomonocytic leukaemia; CMML;  
 KM acute lymphoblastic leukaemia; ALL; APL; NHL; solid tumour;  
 KM papillary thyroid carcinoma; Ewing's sarcoma; melanoma; liposarcoma;  
 KM rhabdomyosarcoma; synovial sarcoma; viral infection.  
 XX  
 OS Homo sapiens.  
 XX  
 PN MO200269900-A2.  
 XX  
 PD 12-SEP-2002.  
 XX  
 PF 01-MAR-2002; 2002MO-US006518.  
 XX  
 PR 01-MAR-2001; 2001US-0272751P.  
 XX  
 PA (CONF-) CONFORMA THERAPEUTICS CORP.  
 XX  
 PI Fritz LC, Burrows FJ;  
 XX  
 DR WPI; 2002-698710/75.  
 XX  
 DR N-PSDB; ABS73274.  
 XX  
 PT Treating genetically-defined disease associated with chromosomal  
 PT aberrations yielding oncogenic fusion proteins, e.g. cell proliferative  
 PT diseases, involves administering an inhibitor of heat shock protein 90.

PS Disclosure; Page 228-229; 389pp; English.

XX The invention describes a method of treating genetically-defined disease

CC associated with chromosomal aberrations yielding oncogenic fusion

CC proteins (I), treating cancerous cells containing (I) in a heterogeneous

CC cell population, treating proliferative diseases associated with mutant

CC protein or cellular protein isoforms (II) dependent on heat shock protein

CC (HSP)-90, or selectively treating cells expressing (II) involving

CC administering HSP90-inhibitor. The method is useful for treating

CC genetically-defined disease with chromosomal aberration yielding

CC oncogenic fusion protein, treating cancerous cells containing fusion

CC protein in heterogeneous cell population, treating proliferative disease

CC (e.g. rheumatoid arthritis or cancer) associated with mutant protein or

CC cellular protein isoform dependent on heat shock protein (HSP)-90 (e.g.

CC p53), or selectively treating cells expressing mutant protein or cellular

CC protein isoform in a patient heterozygous for (II). The method is useful

CC for treating a disease e.g. hematopoietic disorder such as T or B cell

CC lymphoma, chronic myeloid leukaemia (CML), APL, ALL, NHL and CML,

CC or a disease characterized by a solid tumour such as papillary thyroid

CC carcinoma, Ewing's sarcoma, melanoma, liposarcoma, rhabdomyosarcoma and

CC synovial sarcoma. The method is also useful for treating viral

CC infections. This represents a protein encoded by the DNA sequence of a

CC chromosome aberration

XX

XX Sequence 525 AA;

XX

Query Match 45.7%; Score 878; DB 5; Length 525;

Best Local Similarity 52.9%; Pred. No. 5,6e-63; Indels 72; Gaps 13;

Matches 182; Conservative 36; Mismatches 54;

QY 1 GGRGNGSAGRGFGFKPGPMDEGPDLDLGPVDEDESDNSAIYVQGLNDSVTLDDLA 60

DB 248 GGRGNGGS-DRGGFKFGGPRDQSGRHD-----SEGDNSDNTTIFVQGLNDSVTLDDLA 301

QY 61 DFRKQGVVMMKRTQGMPIHITLDKETGPKGADATVSYDPPAKAAVEMFGKDFQS 120

DB 302 DYFKQIGITKTKTKTQPMINLTDTRETGLKGBATVSPDPSAKAAIDWFGKESGN 361

QY 121 KLVNLSARKKPPNMSRGGPLPRGGRGMPPLLRGGPGGPGGPMGRMGGRGDDGFP 180

DB 362 PIVSFAITRADN--RGGNGRGR-----RGGPMRGCGYGG-----GGGSGRGGFP 410

QY 181 PRGPRGRGNPSGGNVORHAGDMQCPNPGCGNFMARTECNQCAKPEGEPLPPFP 240

DB 411 SGG-----GGGGGQQAADWCKNPFCEMNNFSMRNECNQCAKPKDG-----453

QY 241 PGDRGRGGRGGR--GGGGLMNRGGRGGRF--GGGDRGGRGGR--GMRGGRGG 295

DB 454 PG-----GGPGSGHMGNYGDDRRGGGGRGGRGGRGGRGGRGGRGGRGGRG 505

QY 296 RRGPGGPPGLMEQWGRGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGRGGR 338

DB 506 -----GGKMDSRGHRDRRERY 525

RESULT 12

ID ADRI4649 standard; protein; 525 AA.

XX ADRI4649;

XX

XX 21-OCT-2004 (first entry)

XX

DE Human NF-kappaB pathway-associated protein SegID50.

XX NF-kappaB pathway; antiinflammatory; cytosolic; hepatotropic; virucide;

XX antiarthritic; antirheumatic; gastrointestinal; antiasthmatic;

XX antiarteriosclerotic; immunomodulator; cerebroprotective; vasotropic;

XX immunosuppressive; vulnere; gene therapy; immune disorder;

XX inflammatory disorder; NF-kappaB regulation; cancer; aberrant apoptosis;

XX hepatic disorder; Hodgkin's lymphoma; hematopoietic tumour;

XX hyper-IgM syndrome; hypohidrotic ectodermal dysplasia;

XX X-linked anhidrotic ectodermal dysplasia; immunodeficiency;

KW viral infection; HIV-1; HTLV-1; hepatitis B; hepatitis C; EBV; influenza;

KW viral replication; host cell survival; evasion of immune response;

KW rheumatoid arthritis; inflammatory bowel disease; colitis; asthma;

KW atherosclerosis; cachexia; euthyroid sick syndrome; stroke; EAE;

KW autoimmune disorder; hyper immune activity;

KW aberrant acute phase response; hypercongenital condition; birth defect;

KW necrotic lesion; wound; organ transplant rejection;

KW aberrant signal transduction; proliferating disorder; cancer;

KW HIV propagation; human.

OS Homo sapiens.

XX

XX WO2004065577-A2.

XX

XX 05-AUG-2004.

XX

XX 13-JUN-2004; 2004WO-US000798.

XX

XX 14-JAN-2003; 2003US-0440068P.

XX

XX 12-MAY-2003; 2003US-0469757P.

XX (BRIM ) BRISTOL-MYERS SQUIBB CO.

XX

XX Nadler SG, Neubauer MG, Feder JN, Carman J;

PI MPI; 2004-562168/54.

DR N-PSDB; ADRI4648.

XX

XX New isolated polynucleotides and polypeptides associated with NF-kappaB

PT pathway, useful for diagnosing, treating, or preventing disorders or

PT diseases associated with NF-kappaB pathway.

XX

PS Claim 6; SEQ ID NO 650; 237pp; English.

XX

CC This invention relates to the novel association of protein sequences (and

CC the genes which encode them) to the NF-kappaB pathway. The invention may

CC be useful for the production of compounds with an antiinflammatory,

CC cytosolic, hepatotropic, virucide, antiasthmatic, antirheumatic,

CC gastrointestinal-gen, antiasthmatic, antiarteriosclerotic,

CC immunomodulator, cerebroprotective, vasotropic, immunosuppressive or

CC vulnere activity or for gene therapy. The proteins and nucleotides are

CC useful for diagnosing, preventing, treating, or ameliorating conditions

CC or diseases associated with the NF-kappaB pathway. The condition is an

CC immune disorder, an inflammatory disorder, an inflammatory disorder

CC related to aberrant NF-kappaB regulation, cancer, aberrant apoptosis,

CC hepatic disorders, Hodgkin's lymphoma, hematopoietic tumours, hyper-IgM

CC syndromes, hypohidrotic ectodermal dysplasia, X-linked anhidrotic

CC ectodermal dysplasia, immunodeficiency, viral infections, HIV-1, HTLV-1,

CC hepatitis B, hepatitis C, EBV, influenza, viral replication, host cell

CC survival, evasion of immune responses, rheumatoid arthritis, inflammatory

CC bowel disease, colitis, asthma, atherosclerosis, cachexia, euthyroid sick

CC syndrome, stroke, EAE, autoimmune disorders, disorders related to hyper

CC immune activity, disorders related to aberrant acute phase responses,

CC hypercongenital conditions, birth defects, necrotic lesions, wounds,

CC organ transplant rejection, conditions related to organ transplant

CC rejection, disorders related to aberrant signal transduction,

CC proliferating disorders, cancers and HIV propagation in cells infected

CC with other viruses. The present sequence is that of a human protein which

CC is subject to the novel association with the NF-kappaB pathway of the

CC invention. Note: This sequence does not appear in the specification but

CC was obtained by the indexer from Genbank.

XX

XX Sequence 525 AA;

XX

Query Match 45.7%; Score 878; DB 8; Length 525;

Best Local Similarity 52.9%; Pred. No. 5,6e-63;

Matches 182; Conservative 36; Mismatches 54; Indels 72; Gaps 13;

QY 1 GGRGNGSAGRGFGFKPGPMDEGPDLDLGPVDEDESDNSAIYVQGLNDSVTLDDLA 60

DB 248 GGRGNGGS-DRGGFKFGGPRDQSGRHD-----SEGDNSDNTTIFVQGLNDSVTLDDLA 301

QY 61 DFRKQGVVMMKRTQGMPIHITLDKETGPKGADATVSYDPPAKAAVEMFGKDFQS 120

```
DB 302 DYKQIGITKNTKQTPMINTLTDRBTGLKGEATVSPDPSSAALAIWPFQKFSGN 361
QY 121 KLVSLARKKPPMNSMRGGLPPREGGMPPPLRGPGPGPGMGRMGGRGDRGFP 180
DB 362 PIVSFATIRADFN--RGGNNGRG--RGEPWGRGGYGGG--GSGGGGRGFP 410
QY 181 PRGPRGRGNPSSGGVNVQIRADGWOCNPNGCQNPAMRTECNOCAKPRKEGFLPPFP 240
DB 411 SGG-----GGGGGQORADGDMKCPNPTCEMNPFSWNECNOCAKPRPDG----- 453
QY 241 PGDGRGGRGPGMR--GGRGLMDRGPGGMPFG--GGRGDRGPGRGR--GMDRGGPGG 295
DB 454 PG-----GGGGSHMGNGYDDBDRGGRGGYDRGGYRGGGDRGGFRGGGRGDRGFP--- 505
QY 296 RRGCGPGPPGLMEQWGRRGGRGPGKMD--KGEHROERRDRPY 338
DB 506 -----GPGKMDSRGEHROERRERY 525

RESULT 13
AAM78355
ID AAM78355 standard; protein; 526 AA.
XX
AC AAM78355,
XX
DT 06-NOV-2001 (first entry)
XX
DE Human protein SEQ ID NO 1017.
XX
KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorder; arthritis; inflammation.
XX
OS Homo sapiens.
XX
PN WO200157190-A2.
XX
PD 09-AUG-2001.
XX
PF 05-FEB-2001; 2001WO-US004098.
XX
PR 03-FEB-2000; 2000US-00496914.
PR 27-APR-2000; 2000US-00560875.
PR 20-JUN-2000; 2000US-00598075.
PR 19-JUL-2000; 2000US-00620325.
PR 01-SEP-2000; 2000US-00654936.
PR 15-SEP-2000; 2000US-00663561.
PR 20-OCT-2000; 2000US-00693325.
PR 30-NOV-2000; 2000US-00728422.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YF, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y,
PI Ma Y, Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW,
PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;
XX
XX WPI; 2001-476283/51.
XX
DR N-PSDB; AAKS1488.
XX
PT Nucleic acids encoding polypeptides with cytokine-like activities, useful
PT in diagnosis and gene therapy.
XX
PS Claim 20; Page 3253-3254; 6221pp; English.
XX
CC The invention relates to polynucleotides (AAK51456-AAK53435) and the
CC encoded polypeptides (AAM78323-AAM80302) that exhibit activity elating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
```

```
CC activity, tissue growth factor activity, immunomodulatory activity and
CC actinin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation. Note: Records for SEQ ID NO 2110 (AAK52581), 2111
CC (AAK52582) and 3666 (AAM80020) are omitted as the relevant pages from the
CC sequence listing were missing at the time of publication
XX
SQ Sequence 526 AA;
Query Match 45.7%; Score 878; DB 4; Length 526;
Best Local Similarity 52.9%; Pred. No. 5,6e-63;
Matches 182; Conservative 36; Mismatches 54; Indels 72; Gaps 13;

QY 1 GGRGSGSAGRGCGFKPGGPMDEGPDLDLGPVDEDESDNSAIYQGLNSVTLLDLA 60
DB 249 GGRGSGGSG--DRGGFKFGGPPDQGSRDH-----SEQDSDNNNTIFVQIGENVLTESVA 302
QY 61 DPFKQGVVVMNRGTQPMINTLITDKETGPKGDAVTSYEDPPTAKAAVWPFQKFSGN 120
DB 303 DYKQIGITKNTKQTPMINTLTDRBTGLKGEATVSPDPSSAALAIWPFQKFSGN 362
QY 121 KLVSLARKKPPMNSMRGGLPPREGGMPPPLRGPGPGPGMGRMGGRGDRGFP 180
DB 363 PIVSFATIRADFN--RGGNNGRG--RGEPWGRGGYGGG--GSGGGGRGFP 411
QY 181 PRGPRGRGNPSSGGVNVQIRADGWOCNPNGCQNPAMRTECNOCAKPRKEGFLPPFP 240
DB 412 SGG-----GGGGGQORADGDMKCPNPTCEMNPFSWNECNOCAKPRPDG----- 454
QY 241 PGDGRGGRGPGMR--GGRGLMDRGPGGMPFG--GGRGDRGPGRGR--GMDRGGPGG 295
DB 455 PG-----GGGGSHMGNGYDDBDRGGRGGYDRGGYRGGGDRGGFRGGGRGDRGFP--- 506
QY 296 RRGCGPGPPGLMEQWGRRGGRGPGKMD--KGEHROERRDRPY 338
DB 507 -----GPGKMDSRGEHROERRERY 526

RESULT 14
ABG95081
ID ABG95081 standard; protein; 526 AA.
XX
AC ABG95081;
XX
DT 04-DEC-2002 (first entry)
XX
DE Human translocation (12; 16)(q13; p11) protein #2.
XX
KW Chromosome aberration; oncogenic fusion protein; cancer;
KW proliferative disease; cellular protein isoform; heat shock protein 90;
KW HSP-90; rheumatoid arthritis; cancer; haematopoietic disorder;
KW T cell lymphoma; B cell lymphoma; chronic myeloid leukaemia; CML;
KW acute myeloid leukaemia; AML; chronic myelomonocytic leukaemia; CMML;
KW acute lymphoblastic leukaemia; ALL; APL; NHL; solid tumour;
KW papillary thyroid carcinoma; Ewing's sarcoma; melanoma; liposarcoma;
KW rhabdomyosarcoma; synovial sarcoma; viral infection.
XX
XX Homo sapiens.
XX
PN WO200269900-A2.
XX
PD 12-SEP-2002.
XX
PF 01-MAR-2002; 2002WO-US006518.
XX
PR 01-MAR-2001; 2001US-0272751P.
XX
PA (CONF-) CONFORMA THERAPEUTICS CORP.
XX
PI Fritz LC, Burrows FJ;
XX
XX WPI; 2002-698710/75.
XX
DR N-PSDB; ABG73271.
```

XX Treating genetically-defined disease associated with chromosomal  
 PT aberrations yielding oncogenic fusion proteins, e.g. cell proliferative  
 PT diseases, involves administering an inhibitor of heat shock protein 90.  
 PS Disclosure; Page 222-223; 389pp; English.  
 XX  
 CC The invention describes a method of treating genetically-defined disease  
 CC associated with chromosomal aberrations yielding oncogenic fusion  
 CC proteins (I), treating cancerous cells containing (I) in a heterogeneous  
 CC cell population, treating proliferative diseases associated with mutant  
 CC protein or cellular protein isoforms (II) dependent on heat shock protein  
 CC (Hsp)-90, or selectively treating cells expressing (II) involving  
 CC administering HSP90-inhibitor. The method is useful for treating  
 CC genetically-defined disease with chromosomal aberration yielding  
 CC oncogenic fusion protein, treating cancerous cells containing fusion  
 CC protein in heterogeneous cell population, treating proliferative disease  
 CC (e.g. rheumatoid arthritis or cancer) associated with mutant protein or  
 CC cellular protein isoform dependent on heat shock protein (Hsp)-90 (e.g.  
 CC p53), or selectively treating cells expressing mutant protein or cellular  
 CC protein isoform in a patient heterozygous for (II). The method is useful  
 CC for treating a disease e.g. hematopoietic disorder such as T or B cell  
 CC lymphoma, chronic myeloid leukaemia (CML), APL, ALL, NHL and CML,  
 CC or a disease characterised by a solid tumour such as papillary thyroid  
 CC carcinoma, Ewing's sarcoma, melanoma, liposarcoma, rhabdomyosarcoma and  
 CC synovial sarcoma. The method is also useful for treating viral  
 CC infections. This represents a protein encoded by the DNA sequence of a  
 CC chromosome aberration  
 CC  
 SQ Sequence 526 AA;  
 Query Match 45.7%; Score 878; DB 5; Length 526;  
 Best Local Similarity 52.9%; Pred. No. 5.6e-63;  
 Matches 182; Conservative 36; Mismatches 54; Indels 72; Gaps 13;  
 QY 1 GGRGMSAGRGSGFNGKPGMDGPDLDGPPVDDSDNSAIYVGLNDSVTLDDLA 60  
 DB 249 GGRGMSAGRGSGFNGKPGMDGPDLDGPPVDDSDNSAIYVGLNDSVTLDDLA 60  
 QY 61 DFFKQGVVVMKRTQPMIHIYLDKRTGPKDQATVSEDPPTAKAAVEMFGKDFQGS 120  
 DB 303 DFFKQGVVVMKRTQPMIHIYLDKRTGPKDQATVSEDPPTAKAAVEMFGKDFQGS 120  
 QY 121 KATVSLARKKPPMNSRGGLPRBGRMPPLRGSGFGSGPGMGRMGSGGDRGGFP 180  
 DB 363 KATVSLARKKPPMNSRGGLPRBGRMPPLRGSGFGSGPGMGRMGSGGDRGGFP 180  
 QY 181 PRGPRGSRGNPSSGQVQHPADGMOCPNPGCGNPFAMRTGCKAPKEGFLPPFP 240  
 DB 412 PRGPRGSRGNPSSGQVQHPADGMOCPNPGCGNPFAMRTGCKAPKEGFLPPFP 240  
 QY 241 PGSDRGHGFQSGMR-GGRGGLMDRGFGSGMFRG--GGRGDRGGFRGR-GMDRGFGGG 295  
 DB 455 PGSDRGHGFQSGMR-GGRGGLMDRGFGSGMFRG--GGRGDRGGFRGR-GMDRGFGGG 295  
 QY 296 RRGSGPGRPPRLMOMGRRGRGSGRGKMD-KGHNQERDRPY 338  
 DB 507 RRGSGPGRPPRLMOMGRRGRGSGRGKMD-KGHNQERDRPY 338  
 RESULT 15  
 AD126117  
 ID AD126117 standard; protein; 526 AA.  
 XX  
 AC AD126117;  
 XX  
 DT 22-Apr-2004 (first entry)  
 DE Human protein that promotes STAT6 activation #41.  
 XX  
 KM human; signal transducer and activator of transcription 6; STAT6;  
 KM immunogen; STAT6 activation; allergy; inflammation; autoimmune disease;  
 KM diabetes; hyperlipidemia; infection; cancer; Th1 hyperactive disease;

KW rheumatoid arthritis; osteoarthritis; systemic lupus erythematosus;  
 KW sepsis; asthma; allergic rhinitis; ischaemic heart disease;  
 KW subarachnoid haemorrhage; viral hepatitis; AIDS.  
 OS Homo sapiens.  
 PN WO2003104277-A2.  
 PD 18-DEC-2003.  
 PF 05-JUN-2003; 2003WO-JP007123.  
 PR 05-JUN-2002; 2002JP-00164257.  
 PR 06-JUN-2002; 2002US-0385912P.  
 PR 26-DEC-2002; 2002JP-0377326.  
 PR 27-DEC-2002; 2002US-0436467P.  
 PR 15-MAY-2003; 2003JP-00137505.  
 PR 16-MAY-2003; 2003US-0470836P.  
 PA (ASAH ) ASAMI KASRI KK.  
 PI Sugahara T, Matsuda A, Honda G, Muramatsu S, Ishizawa K;  
 DR N-PSDB; AD126116.  
 XX  
 PT New signal transducer and activator of transcription 6 activation  
 PT promoting purified protein, for diagnosing and treating disease  
 PT associated with activation/inhibition of transcription factor e.g.  
 PT diabetes and cancer.  
 PS  
 PS Claim 1; SEQ ID NO 82; 1368pp; English.  
 XX  
 CC The invention relates to a purified protein promoting signal transducer  
 CC and activator of transcription 6 activation (STAT6). The protein is  
 CC useful for the producing an antibody, which involves administering the  
 CC protein or its epitope-bearing fragments to a non-human animal as an  
 CC antigen. The nucleic acid is useful for diagnosing a disease or  
 CC susceptibility to a disease related to expression or activity of the  
 CC protein. A transformant expressing the protein is useful for screening  
 CC compounds which inhibit or promote STAT6 activation. A transformant  
 CC expressing the protein is useful for producing a pharmaceutical  
 CC composition. Compositions, antibodies and antisense molecules are useful  
 CC for the treating a disease associated with STAT6 activation such as  
 CC allergic diseases, inflammation, autoimmune diseases, diabetes,  
 CC hyperlipidemia, infectious disease and cancers. Compositions are useful  
 CC for treating disease associated with STAT6 activation and/or prevention  
 CC of Th1 hyperactive diseases. Compositions are also useful in rheumatoid  
 CC arthritis, osteoarthritis, systemic lupus erythematosus, sepsis, asthma,  
 CC allergic rhinitis, ischaemic heart diseases, subarachnoid haemorrhage,  
 CC viral hepatitis and AIDS. The protein has efficient promoting STAT6  
 CC activity. The protein or nucleic acid is effectively useful for screening  
 CC compounds for treating and preventing disease associated with excessive  
 CC activation or inhibition of STAT6. The present sequence represents the  
 CC amino acid sequence of a human protein which promotes STAT6 activation.  
 XX  
 SQ Sequence 526 AA;  
 Query Match 45.7%; Score 878; DB 8; Length 526;  
 Best Local Similarity 52.9%; Pred. No. 5.6e-63;  
 Matches 182; Conservative 36; Mismatches 54; Indels 72; Gaps 13;  
 QY 1 GGRGMSAGRGSGFNGKPGMDGPDLDGPPVDDSDNSAIYVGLNDSVTLDDLA 60  
 DB 249 GGRGMSAGRGSGFNGKPGMDGPDLDGPPVDDSDNSAIYVGLNDSVTLDDLA 60  
 QY 61 DFFKQGVVVMKRTQPMIHIYLDKRTGPKDQATVSEDPPTAKAAVEMFGKDFQGS 120  
 DB 303 DFFKQGVVVMKRTQPMIHIYLDKRTGPKDQATVSEDPPTAKAAVEMFGKDFQGS 120  
 QY 121 KATVSLARKKPPMNSRGGLPRBGRMPPLRGSGFGSGPGMGRMGSGGDRGGFP 180  
 DB 363 KATVSLARKKPPMNSRGGLPRBGRMPPLRGSGFGSGPGMGRMGSGGDRGGFP 180

```

Qy 181 PRGPRGSRGNPSGGGNVORAGDMOCNPGCGNONFAMRTECHNOCKAPKEGFLPPPP 240
Db 412 SGG-----GGGGGQORAGDMKCPNPTCENNNFSWRNECNQCKAPKPDG----- 454
Qy 241 PGDRGRGGFGGMR-GRGGLMDRGFGGMRG--GRGDRGGFRGGR-GMDRGFGG 295
Db 455 PG-----GGFGGSHMGNGYDDBRGGRGGYDRGGYRGRGGDRGGFRGGGRGGDRGGF--- 506
Qy 296 RRGFGGPPGPELMEQMGRRGRGGPGKMD-KGEHRQERRRPY 338
Db 507 -----GPGKMDSRGEHRQDRRRRPY 526

```

Search completed: February 18, 2005, 15:27:34  
 Job time : 166 secs

**This Page Blank (uspto)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

## OM protein - protein search, using sw model

Run on: February 18, 2005, 15:22:02 / Search time 45 Seconds

(without alignments)  
560.697 Million cell updates/sec

Title: US-10-791-017a-2\_COPY\_319\_656

Perfect score: 1922  
Sequence: 1 GGRGCMGSGRGERGFRKPG.....GGPRKMDKGRHQRDRPY 338Scoring table: BLOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

## Database :

Issued Patents AA:\*  
1: /cgn2\_6/prodata/1/1aa/5A\_COMB.pep:\*  
2: /cgn2\_6/prodata/1/1aa/5B\_COMB.pep:\*  
3: /cgn2\_6/prodata/1/1aa/6A\_COMB.pep:\*  
4: /cgn2\_6/prodata/1/1aa/6B\_COMB.pep:\*  
5: /cgn2\_6/prodata/1/1aa/6CTUS\_COMB.pep:\*  
6: /cgn2\_6/prodata/1/1aa/backfile01.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1922	100.0	656	2	US-08-343-443B-2
2	1922	100.0	656	3	US-09-214-564A-4
3	1922	100.0	656	4	US-09-538-092-1250
4	1810	94.2	591	4	US-09-949-016-10914
5	1810	94.2	591	4	US-09-949-016-10915
6	1337	69.6	306	4	US-09-949-016-7825
7	878	45.7	526	4	US-09-538-092-1080
8	453	23.6	86	2	US-08-343-443B-7
9	303	15.8	1078	3	US-08-963-825-21
10	303	15.8	1078	3	US-09-500-811-21
11	303	15.8	1078	3	US-09-570-573-21
12	303	15.8	1078	3	US-09-548-608-21
13	301	15.7	1057	3	US-08-931-820-4
14	266.5	13.9	955	4	US-09-949-016-8369
15	265	13.8	1603	4	US-09-949-016-6136
16	264.5	13.8	1492	4	US-08-468-996-12
17	264.5	13.8	1065	1	US-08-642-825-72
18	263.5	13.7	1341	1	US-08-963-825-18
19	263.5	13.7	1341	3	US-09-500-811-18
20	263.5	13.7	1341	3	US-09-570-573-18
21	263.5	13.7	1341	3	US-09-548-608-18
22	263.5	13.7	1461	4	US-09-585-887-9
23	263.5	13.7	1461	4	US-09-585-887-9
24	263.5	13.7	1461	4	US-09-585-887-9
25	262	13.6	310	4	US-09-331-347C-21
26	262	13.6	595	3	US-09-219-849-48
27	262	13.6	595	3	US-09-219-849-50

28	262	13.6	822	3	US-09-219-849-49	Sequence 49, Appl
29	261.5	13.6	1806	4	US-09-919-497-56	Sequence 56, Appl
30	260	13.5	633	1	US-08-642-255-73	Sequence 73, Appl
31	259	13.5	1017	4	US-08-468-996-10	Sequence 10, Appl
32	259	13.5	1057	3	US-08-931-820-3	Sequence 1, Appl
33	258.5	13.4	1060	3	US-08-931-820-3	Sequence 3, Appl
34	258.5	13.4	1418	3	US-08-963-825-20	Sequence 20, Appl
35	258.5	13.4	1418	3	US-09-010-999-1	Sequence 1, Appl
36	258.5	13.4	1418	3	US-09-500-811-20	Sequence 20, Appl
37	258.5	13.4	1418	3	US-09-570-573-20	Sequence 20, Appl
38	258.5	13.4	1418	3	US-09-548-608-20	Sequence 20, Appl
39	258	13.4	689	4	US-09-949-016-11276	Sequence 11276, A
40	258	13.4	1609	4	US-09-949-016-5883	Sequence 5883, Ap
41	258	13.4	1670	4	US-09-949-016-5883	Sequence 8890, Ap
42	257	13.4	599	4	US-09-949-016-8890	Sequence 417, App
43	253	13.2	274	4	US-09-976-594-417	Sequence 12, Appl
44	253	13.2	1442	2	US-08-316-650-12	
45	253	13.2	1442	5	PCT-US95-02251-12	

## ALIGNMENTS

RESULT 1  
US-08-343-443B-2  
Sequence 2, Application US/08343443B  
Patent No. 5968734  
GENERAL INFORMATION:  
APPLICANT: Aurias, Alain  
APPLICANT: Delattre, Olivier  
APPLICANT: Desmaret, Chantal  
APPLICANT: Meiot, Thomas  
APPLICANT: Peter, Martine  
APPLICANT: Ploougascel, Beatrice  
APPLICANT: Thomas, Jessica  
TITLE OF INVENTION: NUCLEIC ACID CORRESPONDING TO A GENE OF  
TITLE OF INVENTION: CHROMOSOME 22 INVOLVED IN RECURRENT CHROMOSOMAL  
TITLE OF INVENTION: TRANSLATIONS ASSOCIATED WITH THE DEVELOPMENT OF CANCEROUS  
TITLE OF INVENTION: TUMORS, AND NUCLEIC ACIDS OF FUSION RESULTING FROM SAID  
NUMBER OF SEQUENCES: 129  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Weiser & Associates  
STREET: 230 South Fifteenth Street  
CITY: Philadelphia  
STATE: PA  
COUNTRY: USA  
ZIP: 19102  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: AEDIT 1.0 DOS text editor  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/343,443B  
FILING DATE: 18-NOV-1994  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/FR93/00494  
FILING DATE: 19-MAY-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: FR 92/06123  
FILING DATE: 20-MAY-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Weiser, Gerard J.  
REGISTRATION NUMBER: 19,763  
REFERENCE/DOCKET NUMBER: 989.6121P  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 215-875-8383  
TELEFAX: 215-875-8394  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:

```

; LENGTH: 656 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-343-443B-2

```

Query Match	100.0%	Score 1922;	DB 2;	length 656;
Best Local Similarity	100.0%	Pred. No. 1.7e-150;		
Matches 338;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0

QY 1 GGRGMSGASBERGGFNKPGGEMDEGDDLGLPPVDEDEDSNSAIYVQGLNDSVTLLDILA 60  
DB 319 GGRGMSGASBERGGFNKPGGEMDEGDDLGLPPVDEDEDSNSAIYVQGLNDSVTLLDILA 378

QY 61 DFFKCGVYMNKRIGQPMHIYLDKETGKKGDAVTSYEBPPTAKAAVENEDEGKDFQS 120  
DB 379 DFFKCGVYMNKRIGQPMHIYLDKETGKKGDAVTSYEBPPTAKAAVENEDEGKDFQS 438

**Oy**

**Db**

**KLKTSLARKKP PMSMRGGLP PRGRGMPP LRGCGCG PGPGG PMRMGGRR GDRGGFP 180**

**KLKTSLARKKP PMSMRGGLP PRGRGMPP LRGCGCG PGPGG PMRMGGRR GDRGGFP 498**

**Oy** 181 PRCPGRSGRNPSGGGNVQHRA GDMQCPNPGCNQNFA M R T E N Q C K A P K B E F L P P P P P 240  
**Ddb** 499 PRCPGSRGNPSSGGGNVQHRAGD MQCPNPGCNQNFA M R T E N Q C K A P K B E F L P P P P P 558	

**Oy**      241 PGDGRGCGPGMCGRGGLMDRGGCGMPFRGGRGGRGCGFPFGRCMDRGCFPGGRRGCP 300  
|||||  
|||

**Db**      559 PGGDRGRCGPCCGMRCGRGGLMDRGGCGMPFRGGRGGRGCGFPFGRCMDRGCFGGRRGCP 618  
|||||  
|||

```
OY      301 GSPGPELMEQMGRRGGSGPGKMDKGHEHRDRDREY   338  
        |||||||  
Db      619 GGPPPLMEQMGRRCRGSGPGKMDKGHEHRRDRDREY   656
```

RESULT 2  
US-09-214-564A-4

```

; sequence # Application 08/052152076
; Patent No. 6150515
; GENERAL INFORMATION:
; APPLICANT: Sharp, Phillip A.

```

```

; AFFILIANT: ZHOU, YIANG
; TITLE OF INVENTION: TAT-SF: Cofactor For Stimulation Of Transcriptional
; TITLE OF INVENTION: Elongation By HIV-1 TAT
; FILE REFERENCE: M0656/7042

```

;  
 ; CURRENT APPLICATION NUMBER: US 05/ 214,567A  
 ;  
 ; CURRENT FILING DATE: 1999-08-18  
 ;  
 ; PRIOR APPLICATION NUMBER: US 60/021,218  
 ;  
 ; PRIOR FILING DATE: 1996-07-03  
 ;

; PRIOR APPLICATION NUMBER: US 60/033,152  
 ; PRIOR FILING DATE: 1996-12-13  
 ; PRIOR APPLICATION NUMBER: PCT/US97/11713  
 ; PRIOR FILING DATE: 1997-07-03

```

; NUMBER OF SEQ ID NOS: 5
;
; SOFTWARE: FastSeq for Windows Version 3.0
;
; SEQ ID NO 4
;
; LENGTH: 656

```

```

; TYPE: PRI
; ORGANISM: Homo sapiens
US-09-214-564A-4

```

Query Match	100.0%;	Score 1922;	DB 3;	Length 656;
Best Local Similarity	100.0%;	Pred. No. 1.7e-150;		
Matches 338;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0

QY 1 GGRGNGSAGERGGFNKPGGPMDEGPDLDPVDPDSDNSAIYVGINDSTLTLDLIA 60  
DB 319 GGRGNGSAGERGGFNKPGGPMDEGPDLDPVDPDSDNSAIYVGINDSTLTLDLIA 378

QY 61 DFFKQGVVKNKRRGTGPMIHLYLDKETGPKKDATVSYEDPPTAKAAVEMFGKDFQGS 120

DB 379 DFFKQGVVKNKRRGTGPMIHLYLDKETGPKKDATVSYEDPPTAKAAVEMFGKDFQGS 438

Qy	121	KLVSLIARKKPPMNSMRGLPREERGMPLRGGPGSGPGGPMRMGARGDGRGFP	180
Db	439	KLVSLIARKKPPMNSMRGLPREERGMPLRGGPGSGPGGPMRMGARGDGRGFP	498
Qy	181	PRGPRSGRNPSSGGGNVHRAGDMQCPNPCCGNONFAMRTETCNCCKAPKEEGFLPPFP	240
Db	499	PRGPRSGRNPSSGGGNVHRAGDMQCPNPCCGNONFAMRTETCNCCKAPKEEGFLPPFP	558
Qy	241	PGDGRGRGPGMRRGGGLMDRGGPGGMFRGSGRGDRGFRGRRGMDRGGFGGRRGCGP	300
Db	559	PGDGRGRGPGMRRGGGLMDRGGPGGMFRGSGRGDRGFRGRRGMDRGGFGGRRGCGP	618
Qy	301	GGPGRFLMEOMGRRGRRGCGPKMDKGEHRDERDRPY	338
Db	619	GGPGRFLMEOMGRRGRRGCGPKMDKGEHRDERDRPY	656

RESULT 3  
US-09-538-092-1250  
; Sequence 1250, Application US/09538092

```

; GENERAL INFORMATION:
;
; APPLICANT: Giot, Loic
; APPLICANT: Mansfield, Traci A.

```

FILE REFERENCE: 15966-542  
CURRENT APPLICATION NUMBER: US/09/538,092  
CURRENT FILING DATE: 2000-03-29

PRIOR FILING DATE: 1999-04-01  
PRIOR APPLICATION NUMBER: 60/178,965  
PRIOR FILING DATE: 2000-02-01

```

; NUMBER OF SEQ ID NOS: 1250
; SOFTWARE: CuratPatSeqFormatter Version 0.9
; SEQ ID NO 1250
; LENGTH: 656

```

```

; LIFE: FRI
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature

```

```

? LOCATION: (U)...(U)
? OTHER INFORMATION: Polypeptide Accession Number Q01844
US-09-538-092-1250

```

Query Match	100.0%;	Score 1922;	DB 4;	Length 656;
Best local Similarity	100.0%;	Pred. No. 1.7e-150;		
Matches 338;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0

**QY**      1 GGRGCGSAGRRGGFNKPGGPMDEGDLDLGPVPVDSNSAIYQGINSVLINDPRA 80  
**Db**    319 GGRGCGSAGRRGGFNKPGGPMDEGDLDLGPVPVDSNSAIYQGINSVTLLDPLA 378

QY 61 DFFKCGVKKKKRIGQPMHIYLDKETGKPGDATSYEDPTAKAAVEMFEGSKD FGS 120  
Db 379 DFFKCGVKKKKRIGQPMHIYLDKETGKPGDATSYEDPTAKAAVEMFEGSKD FGS 438

Qy 121 KLIVSLARKKCPMNINRGLDPRREGGMPDPIRGSGPGGSGPMGMGMGRGGDRGGP 18

Db 439 KLIVSLARKKCPMNINRGLDPRREGGMPDPIRGSGPGGSGPMGMGMGRGGDRGGP 496

QY 181 PRGPRGSRGNPSGGNNVQHRAGDWQCPNPGCGNQNFAMRTENQCKAPKEGFLPPPPP 241

DB 499 PRGPRGSRGNPSGGNNVQHRAGDWQCPNPGCGNQNFAMRTENQCKAPKEGFLPPPPP 550

Qy 241 PGGDRRGPGGMRGGGGLNDRGCGMFRGGRGGRGFRGGGRGGP 300  
Db 559 PGGDRRGPGGMRGGGGLNDRGCGMFRGGRGGRGFRGGGRGGP 610

QY 301 GSPPGPLMEQMGRRGSGRGGPGKMDKGHEHQDRDRPY 338  
|||  
DB 619 GSPPGPLMEQMGRRGSGRGGPGKMDKGHEHQDRDRPY 656

```
RESULT 4
US-09-949-016-10914
; Sequence 10914, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10914
; LENGTH: 591
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-10914

Query Match
Best Local Similarity 100.0%; Pred. No. 2.6e-141; Length 591;
Matches 318; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 21 PMDEGPDLDGPPVDEDESDNSAIYVGLNDSVTLLDLADFPKQCGVVMNKRRTGQPMI 80
DB 274 PMDEGPDLDGPPVDEDESDNSAIYVGLNDSVTLLDLADFPKQCGVVMNKRRTGQPMI 333
QY 81 HTYLDKETGPKGDATVSYEDPPTAAVAWEPFGKDFQSGKLKVLARKKPPMNSMRGGL 140
DB 334 HTYLDKETGPKGDATVSYEDPPTAAVAWEPFGKDFQSGKLKVLARKKPPMNSMRGGL 393
QY 141 PPRGGMPPPLARGGGPGGPGMGRMGGRGDDGPPRRPRGRSGRNPSSGGVQHR 200
DB 394 PPRGGMPPPLARGGGPGGPGGPGMGRMGGRGDDGPPRRPRGRSGRNPSSGGVQHR 453
QY 201 AGDMQCPNPGCGNQNFAMRTECNQCKAPKEGFLPPFPFGDGRGPGGMRGGRGGL 260
DB 454 AGDMQCPNPGCGNQNFAMRTECNQCKAPKEGFLPPFPFGDGRGPGGMRGGRGGL 513
QY 261 MDRGPGGMRGGRGGRGDRGFRGGRGMDRGFGGRRGGGPGPPLMEQMGRRGGRG 320
DB 514 MDRGPGGMRGGRGGRGDRGFRGGRGMDRGFGGRRGGGPGPPLMEQMGRRGGRG 573
QY 321 PGKMDKEHROERDRPY 338
DB 574 PGKMDKEHROERDRPY 591

RESULT 5
US-09-949-016-10915
; Sequence 10915, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
```

```
; SEQ ID NO 10915
; LENGTH: 591
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-10915

Query Match
Best Local Similarity 100.0%; Pred. No. 2.6e-141; Length 591;
Matches 318; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 21 PMDEGPDLDGPPVDEDESDNSAIYVGLNDSVTLLDLADFPKQCGVVMNKRRTGQPMI 80
DB 274 PMDEGPDLDGPPVDEDESDNSAIYVGLNDSVTLLDLADFPKQCGVVMNKRRTGQPMI 333
QY 81 HTYLDKETGPKGDATVSYEDPPTAAVAWEPFGKDFQSGKLKVLARKKPPMNSMRGGL 140
DB 334 HTYLDKETGPKGDATVSYEDPPTAAVAWEPFGKDFQSGKLKVLARKKPPMNSMRGGL 393
QY 141 PPRGGMPPPLARGGGPGGPGMGRMGGRGDDGPPRRPRGRSGRNPSSGGVQHR 200
DB 394 PPRGGMPPPLARGGGPGGPGGPGMGRMGGRGDDGPPRRPRGRSGRNPSSGGVQHR 453
QY 201 AGDMQCPNPGCGNQNFAMRTECNQCKAPKEGFLPPFPFGDGRGPGGMRGGRGGL 260
DB 454 AGDMQCPNPGCGNQNFAMRTECNQCKAPKEGFLPPFPFGDGRGPGGMRGGRGGL 513
QY 261 MDRGPGGMRGGRGGRGDRGFRGGRGMDRGFGGRRGGGPGPPLMEQMGRRGGRG 320
DB 514 MDRGPGGMRGGRGGRGDRGFRGGRGMDRGFGGRRGGGPGPPLMEQMGRRGGRG 573
QY 321 PGKMDKEHROERDRPY 338
DB 574 PGKMDKEHROERDRPY 591
```

```
RESULT 6
US-09-949-016-7825
; Sequence 7825, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7825
; LENGTH: 306
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-7825

Query Match
Best Local Similarity 78.3%; Pred. No. 1.3e-102; Length 306;
Matches 246; Conservative 13; Mismatches 37; Indels 18; Gaps 2;

QY 21 PMDEGPDLDGPPVDEDESDNSAIYVGLNDSVTLLDLADFPKQCGVVMNKRRTGQPMI 80
DB 10 PMDEGPDLDGPPVDEDESDNSAIYVGLNDSVTLLDLADFPKQCGVVMNKRRTGQPMI 69
QY 81 HTYLDKETGPKGDATVSYEDPPTAAVAWEPFGKDFQSGKLKVLARKKPPMNSMRGGL 140
DB 70 HTYLDKETGPKGDATVSYEDPPTAAVAWEPFGKDFQSGKLKVLARKKPPMNSMRGGL 129
QY 141 PPRGGMPPPLARGGGPGGPGMGRMGGRGDDGPPRRPRGRSGRNPSSGGVQHR 200
```

Db 130 PPRGGWPPPLCGGPGGPGSGPGMGMGGRGDRGGLPPRPGQSRGNTSGGVVQ 189  
QY 201 AGMOCNPGCCGNQNMAMRTCCNOCAPRREGLPPPPPPGGDRGRGGRGGL 260  
Db 190 AGROCPNPGCCGNQNMAMRTCCNOCAPRREGLPPPPPPGGDRGRGGRGGL 249  
QY 261 MDGPGGMPGRRGGDRGGDRGGDRGGDRGGDRGGDRGGDRGGDRGGDRGG 320  
Db 250 MDGPGGMPGRRGGDRGGDRGGDRGGDRGGDRGGDRGGDRGGDRGGDRGG 291  
QY 321 PGMDKGEHRQERR 334  
Db 292 PRFDVTNGRKKRR 305

## RESULT 7

US-09-538-092-1080  
; Sequence 1080, Application US/09538092  
; Patent No. 6753314  
; GENERAL INFORMATION:  
; APPLICANT: Glot, Loic  
; APPLICANT: Mansfield, Traci A.  
; TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same  
; FILE REFERENCE: 15966-542  
; CURRENT APPLICATION NUMBER: US/09/538,092  
; PRIOR FILING DATE: 2000-03-29  
; PRIOR APPLICATION NUMBER: 60/127,352  
; PRIOR FILING DATE: 1999-04-01  
; PRIOR APPLICATION NUMBER: 60/178,965  
; PRIOR FILING DATE: 2000-02-01  
; NUMBER OF SEQ ID NOS: 1387  
; SOFTWARE: CuratSeqFormatter Version 0.9  
; SEQ ID NO 1080  
; LENGTH: 526  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (0)...(0)  
; OTHER INFORMATION: Polypeptide Accession Number P35637  
US-09-538-092-1080

## Query Match

45.7%; Score 878; DB 4; Length 526;  
Best Local Similarity 52.9%; Pred. No. 1.7e-64; Indels 72; Gaps 13;  
Matches 182; Conservative 36; Mismatches 54; Indels 72; Gaps 13;

QY 1 GGRGMSAGRRGGGPNRGGPMDGPDLDGPPVDPEDSDNSAIYVQGLNDSVTLDDL 60  
Db 249 GGRGMSAGRRGGGPNRGGPMDGPDLDGPPVDPEDSDNSAIYVQGLNDSVTLDDL 302  
QY 61 DFRKQCCVVMKRTGQPMTHIYLDKGTGPKGDATVSYDDPTAKAAVWFGKDFQGS 120  
Db 303 DFRKQCCVVMKRTGQPMTHIYLDKGTGPKGDATVSYDDPTAKAAVWFGKDFQGS 362  
QY 121 KLVSLARKKPKPMNSMGGLPPRGRGMPPLRGSGPGSGPGSGPGSGPGSGPGSG 180  
Db 363 KLVSLARKKPKPMNSMGGLPPRGRGMPPLRGSGPGSGPGSGPGSGPGSGPGSG 411  
QY 181 PRGPGSGRGNPGSGGNVQHRAGDMQCPNPGCNQNFAMRTCCNOCAPRREGLPPPP 240  
Db 412 SGG-----GGGGGQQAAGDMKCPNPGCNQNFAMRTCCNOCAPRREGLPPPP 454  
QY 241 PGDRGRGPGGGR-GRGGGLMDRGPGGMRG--GRGDRGGFRGR-GMDRGFGGG 295  
Db 455 PG-----GGPGSGHMGNYGDRGRGGRGGRGGRGGRGGRGGRGGRGGRGGRG 506  
QY 296 RRGPGGPPGRLMEQMGRRGRGRGGRGMD-KGHRGRDRRY 338  
Db 507 -----GPGKMSRGRHRRDRRY 526

## RESULT 8

US-08-343-443B-7  
; Sequence 7, Application US/08343443B  
; Patent No. 5968734

; GENERAL INFORMATION:  
; APPLICANT: Aurias, Alain  
; APPLICANT: Delattre, Olivier  
; APPLICANT: Desmarte, Chantal  
; APPLICANT: Meiot, Thomas  
; APPLICANT: Peter, Martine  
; APPLICANT: Ploougaestel, Beatrice  
; APPLICANT: Thomas, Gilles  
; APPLICANT: Zucman, Jessica  
; TITLE OF INVENTION: NUCLEIC ACID CORRESPONDING TO A GENE OF  
; TITLE OF INVENTION: CHROMOSOME 22 INVOLVED IN RECURRENT CHROMOSOMAL  
; TITLE OF INVENTION: TRANSLATIONS ASSOCIATED WITH THE DEVELOPMENT OF CANCEROUS  
; TITLE OF INVENTION: TUMORS, AND NUCLEIC ACIDS OF FUSION RESULTING FROM SAID  
; NUMBER OF SEQUENCES: 129  
; CORRESPONDENCE ADDRESS:  
; ADDRESSER: Weiser & Associates  
; STREET: 230 South Fifteenth Street  
; CITY: Philadelphia  
; STATE: PA  
; COUNTRY: USA  
; ZIP: 19102

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: AEDIT 1.0 DOS text editor  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/343,443B  
FILING DATE: 18-NOV-1994  
CLASSIFICATION: 514

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/FR93/00494  
FILING DATE: 19-MAY-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: FR 92/06123  
FILING DATE: 20-MAY-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Weiser, Gerard J.  
REGISTRATION NUMBER: 19,763  
REFERENCE/DOCKET NUMBER: 989.6121P  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 215-875-8383  
TELEFAX: 215-875-8394

INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 86 amino acids  
TYPE: amino acid  
TOPOLOGY: linear

US-08-343-443B-7

## Query Match

23.6%; Score 453; DB 2; Length 86;  
Best Local Similarity 100.0%; Pred. No. 2.5e-30;  
Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 45 IYVQGLNDSVTLDDLADFRKQCCVVMKRTGQPMTHIYLDKGTGPKGDATVSYDDPT 104  
Db 1 IYVQGLNDSVTLDDLADFRKQCCVVMKRTGQPMTHIYLDKGTGPKGDATVSYDDPT 60  
QY 105 AKAAVWFGKDFQGSKLVSLARKK 130  
Db 61 AKAAVWFGKDFQGSKLVSLARKK 86

## RESULT 9

US-08-963-825-21  
; Sequence 21, Application US/08963825  
; Patent No. 6110689  
; GENERAL INFORMATION:  
; APPLICANT: Oviatt, Per

```

Cy      310 QMGRRGGRGPG-KMDXGE 328
Db      3579 GMPERGGLSPGPKDKGE 598

RESULT 10
US-09-500-811-21
Sequence 21, Application US/09500811
Patent No. 632314
GENERAL INFORMATION:
APPLICANT: Qvist, Per
INVENTOR: Bonde, Martin
TITLE OF INVENTION: A Method for Assaying Collagen Fragments
in Body Fluids, A Test Kit and Means for Carrying Out the
Method and Use of the Method to Diagnose the Presence of
TITLE OF INVENTION: Disorders Associated with the Metabolism of
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSER: Darby & Darby PC
STREET: 805 Third Avenue
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10022
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/500,811
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/187,319
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Gogoris, Adda C
REGISTRATION NUMBER: 29,714
REFERENCE/DOCKET NUMBER: 4305/08701
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-527-7700
TELEFAX: 212-753-6237
TELEX: 236687
INFORMATION FOR SEQ ID NO: 21:
SEQUENCE CHARACTERISTICS:
LENGTH: 1078 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
IMMEDIATE SOURCE:
CLONE: COLLAGEN ALPHA 1 (III)
US-09-500-811-21

Query Match          15.8%; Score 303; DB 3; Length 1078;
Best Local Similarity 30.3%; Pred. No. 1e-16;
Matches 115; Conservative 15; Mismatches 126; Indels 124; Gaps 16;

Cy      1 GGRGNGSAGERGFGNKPQPMDEGPDLGLPVPDPDSDNSALYVGLNDSTLDDLA 60
       ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     291 GAKEGEPGRGERGKREAGIPGVPGAEBDGKDSFGDPGANGLPGAAGERGALGS----- 343

Cy      61 DFKQGCGVVAKMKRKTQGMPIHIYLDKSGKPKGDATVSYEDPPTKAAVEWFDGKDFGS 120
       : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db     344 -----RGAPGNIGIEKGKPAGE-----361

Cy    121 KLVSLARKKKPPNNISNRG--GLPPREGRMPLRLGRCGGGPGPG-----GPMGRWG--G 171
       : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db     362 -----RAFPBPAGRGGAAGFPGDVPDGFGRGMRGMTFSFGPSDGRKPPPSQSBSG 414

Cy    172 RGGDRGRFPPRPGRSGR-----NPSGGNVGHRAADMQCENPCGNGNFAMTECNCK 226

```

Db 415 RRPFG---PSGRGPGVWGFPGPKNDGAPKNGERGP-GRPGG-----459  
Qy 227 APRRGFLPPPP-----PGDRGRGPGGMRG-----GRGLMDR-----263  
Db 460 PPKNGEYGPQPGPTGPGDKDTGPRPGQLGGLPTGGP PGSENEKGPBGKGEAG 519  
Qy 264 ---GSPGMRGPGGGRG-----GFRGGRMDRGPGGGRGPGGPG-----PLMB 309  
Db 520 APAPGKGDAAPGGEGRPGGLAGAPGLRGAG-PPGPGGKGAAGPPGPGAAGTGLQ 578  
Qy 310 QMGRRGRGPGG-KMDKGE 328  
Db 579 GMPGERGGLGSPGPKDKGE 598

RESULT 11  
US-09-570-573-21  
Sequence 21, Application US/09570573  
Patent No. 634361  
GENERAL INFORMATION:  
APPLICANT: Qvist, Per  
TITLE OF INVENTION: A Method for Assaying Collagen Fragments  
TITLE OF INVENTION: In Body Fluids, A Test Kit and Means for Carrying Out the  
TITLE OF INVENTION: Method and Use of the Method to Diagnose the Presence of  
NUMBER OF INVENTIONS: 21  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Darby & Darby PC  
STREET: 805 Third Avenue  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10022  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/570,573  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/187,319  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Gogoris, Adda C  
REGISTRATION NUMBER: 29,714  
REFERENCE/DOCKET NUMBER: 4305/08701  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-527-7700  
TELEFAX: 212-753-6237  
TELEX: 236687  
INFORMATION FOR SEQ ID NO: 21:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1078 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
ORIGINAL SOURCE:  
ORGANISM: Homo sapiens  
IMMEDIATE SOURCE:  
CLONE: COLLAGEN ALPHA 1 (III)  
US-09-570-573-21

Query Match 15.84; Score 303; DB 3; Length 1078;  
Best Local Similarity 30.34; Pred. No. 1e-16;  
Matches 115; Conservative 15; Mismatches 126; Indels 124; Gaps 16;  
1 GGRGMSABRRGFNKGPMDEGPPVDPDSDNSAIYVGINDSVTLDDLA 60

Db 291 GAKGEPGRGERGEAGIDGVPKAKGEDKDGSPDGPAGNGLPGAAERGLGS-----343  
Qy 61 DFFKCGGVKMKKRTGQPMIHTYLDKENGKPGDATTSEDEPTLKAATWFTDKDFGS 120  
Db 344 -----RGPAGNGIPEEKPAGE-----361  
Qy 121 KLKVLARKKPPMNSMRG--GLPPEGRGMPPRLRGPGGPGPG-----GMWRMG--G 171  
Db 362 -----RAPAPAGRGAAGERGRDVGPCGGMGMPSPPGPGSDGKRPFGSGGEG 414  
Qy 172 RGDRGRGPPPRGPRGSG-----NPSGGNVHRAADMOCNPPGCCNQNPAFTTECNQK 226  
Db 415 RRPFG---PSGRGPGVWGFPGPKNDGAPKNGERGP-GRPGG-----459  
Qy 227 APRRGFLPPPP-----PGDRGRGPGGMRG-----GRGLMDR-----263  
Db 460 PPKNGEYGPQPGPTGPGDKDTGPRPGQLGGLPTGGP PGSENEKGPBGKGEAG 519  
Qy 264 ---GSPGMRGPGGGRG-----GFRGGRMDRGPGGGRGPGGPG-----PLMB 309  
Db 520 APAPGKGDAAPGGEGRPGGLAGAPGLRGAG-PPGPGGKGAAGPPGPGAAGTGLQ 578  
Qy 310 QMGRRGRGPGG-KMDKGE 328  
Db 579 GMPGERGGLGSPGPKDKGE 598

RESULT 12  
US-09-548-608-21  
Sequence 21, Application US/09548608  
Patent No. 635442  
GENERAL INFORMATION:  
APPLICANT: Qvist, Per  
TITLE OF INVENTION: A Method for Assaying Collagen Fragments  
TITLE OF INVENTION: In Body Fluids, A Test Kit and Means for Carrying Out the  
TITLE OF INVENTION: Method and Use of the Method to Diagnose the Presence of  
NUMBER OF INVENTIONS: 21  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Darby & Darby PC  
STREET: 805 Third Avenue  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10022  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/548,608  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/187,319  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Gogoris, Adda C  
REGISTRATION NUMBER: 29,714  
REFERENCE/DOCKET NUMBER: 4305/08701  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-527-7700  
TELEFAX: 212-753-6237  
TELEX: 236687  
INFORMATION FOR SEQ ID NO: 21:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1078 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
ORIGINAL SOURCE:



ORGANISM: Homo sapiens  
IMMEDIATE SOURCE:  
CLONE: COLLAGEN ALPHA 1 (III)  
US-09-548-608-21

Query Match 15.8%; Score 303; DB 3; Length 1078;  
Best Local Similarity 30.3%; Pred. No. 1e-16;  
Matches 115; Conservative 15; Mismatches 126; Indels 124; Gaps 16;

QY 1 GGGGMSAGRGGFKKPGGPMDEGPDLDGPPVDPEEDSDNSAIYQGLNDVTLDDL 60  
DB 291 GAGGEPGPRGEBAGIPGVGAKGEGDKGSPEDPANGLPGAAGERGALGS----- 343  
QY 61 DFFKGGVVMNKRQTGPMHIYLDKTKPKKGDAIVSYEDPPTAKAAVEMFGDKDFGS 120  
DB 344 -----RGDAEPNGIPGKGPAGE----- 361  
QY 121 KLVSLARKKPPNNMNG--GLPREGRGMPPLRGPGGPGPG--GPMGRMG--G 171  
DB 362 -----RGAPGAPGAGGAEPRDGVPGGPKRMGPGSPGSGDKKPGPSQGESG 414  
QY 172 RGDRGPFPPRGPSRG-----NPSGGANVQHRADMGCPNPGCCNQNFAMTECNQCK 226  
DB 415 RGPSPG---PSGPRGPGVWGFPGPKNDGAPKNGERGGPG--GPGPG----- 459  
QY 227 APKPEGLPPPPPP---PGDNRGGPGGMRG-----GRGGLMDR----- 263  
DB 460 PPKNGETGPGPPGPTGPGDGKDTGPRPGQLGLPTGTGPGGENKKEGEPKGDAG 519  
QY 264 --GGPGMFRGGGDRG-----GFRGGRGMDRGFGGRRGGPGPG---PLME 309  
DB 520 APGAPGKGDAGAPGEGPRGLAGAPGLRGAG--PFGPEGKKAAGPFGPGAAGTPTGLQ 578  
QY 310 QMGRRGGRGPG--KMDXGE 328  
DB 579 GMPERGGLGSPGPKDKGE 598

## RESULT 13

US-08-931-820-4  
Sequence 4, Application US/08931820  
Patent No. 6010863  
GENERAL INFORMATION:  
APPLICANT:  
TITLE OF INVENTION: Assay for collagen degradation  
NUMBER OF SEQUENCES: 4  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
OPERATING SYSTEM: IBM PC compatible  
SOFTWARE: Patent Release #1.0, Version #1.25 (BPO)  
CURRENT APPLICATION DATA:  
FILING DATE:  
CLASSIFICATION: 435  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: EP 96202596.1  
FILING DATE:  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1057 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
ORIGINAL SOURCE:  
ORGANISM: Homo sapiens  
TISSUE TYPE: Collagen type III  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 1055  
OTHER INFORMATION: /label= Modified

OTHER INFORMATION: /note= "Ala may be Pro"  
US-08-931-820-4

Query Match 15.7%; Score 301; DB 3; Length 1057;  
Best Local Similarity 30.5%; Pred. No. 1.5e-16;  
Matches 116; Conservative 15; Mismatches 125; Indels 124; Gaps 18;

QY 1 GGGGMSAGRGGFKKPGGPMDEGPDLDGPPVDPEEDSDNSAIYQGLNDVTLDDL 60  
DB 290 GAGGEPGPRGEBAGIPGVGAKGEGDKGSPGEP-----GAN----- 328  
QY 61 DFFKGGVVMNKRQTGPMHIYLDKTKPKKGDAIVSYEDPPTAKAAVEMFGDKDFGS 120  
DB 329 -----GLPDAEBRGAHGR-----GPAEPNGIPGKGPAGE----- 360  
QY 121 KLVSLARKKPPNNMNG--GLPREGRGMPPLRGPGGPGPG--GPMGRMG--G 171  
DB 361 -----RGAPGAPGAGGAEPRDGVPGGPKRMGPGSPGSGDKKPGPSQGESG 413  
QY 172 RGDRGPFPPRGPSRG-----NPSGGANVQHRADMGCPNPGCCNQNFAMTECNQCK 226  
DB 414 RGPSPG---PSGPRGPGVWGFPGPKNDGAPKNGERGGPG--GPGPG----- 458  
QY 227 APKPEGLPPPPPP---PGDNRGGPGGMRG-----GRGGLMDR----- 263  
DB 459 PPKNGETGPGPPGPTGPGDGKDTGPRPGQLGLPTGTGPGGENKKEGEPKGDAG 518  
QY 264 --GGPGMFRGGGDRG-----GFRGGRGMDRGFGGRRGGPGPG---PLME 309  
DB 519 APGAPGKGDAGAPGEGPRGLAGAPGLRGAG--PFGPEGKKAAGPFGPGAAGTPTGLQ 577  
QY 310 QMGRRGGRGPG--KMDXGE 328  
DB 578 GMPERGGLGSPGPKDKGE 597

## RESULT 14

US-09-949-016-8369  
Sequence 8369, Application US/09949016  
Patent No. 6812339  
GENERAL INFORMATION:  
APPLICANT: VENTER, J. Craig et al.  
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
FILE REFERENCE: CL001307  
CURRENT APPLICATION NUMBER: US/09/949,016  
PRIORITY FILING DATE: 2000-04-14  
PRIORITY APPLICATION NUMBER: 60/241,755  
PRIORITY FILING DATE: 2000-10-20  
PRIORITY APPLICATION NUMBER: 60/237,768  
PRIORITY FILING DATE: 2000-10-03  
PRIORITY APPLICATION NUMBER: 60/231,498  
PRIORITY FILING DATE: 2000-09-08  
NUMBER OF SEQ ID NOS: 207012  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 8369  
LENGTH: 955  
TYPE: PRT  
ORGANISM: Human  
US-09-949-016-8369

Query Match 13.9%; Score 266.5; DB 4; Length 955;  
Best Local Similarity 30.8%; Pred. No. 9.2e-14;  
Matches 123; Conservative 19; Mismatches 143; Indels 115; Gaps 22;

QY 1 GGRG-----MGSAGERGGRNKRGG---PMDEGPDLDGPP-VPPDED 39  
DB 555 GSGGSPDGAGSKLPPVIANIMGAGKGGPGGGINVQELITSTMGSPNSHPSR- 613  
QY 40 SDSAIVQGLNDVTLDDLADFFKGGVVMNKRQTGPMHIYLDKTKGPKGDATVSG 99  
DB 614 -----ELKQPSYSDIKQM-LVPHGLGPGFIANGPFGGSGGPRG-----NQ 656

QY 100 EDEPTAKAAVEMFQDFOGSKLVSLARKKPPMNSMRGSLPPREGRMPPLRGPP--G 157  
DB 657 HFPFGGPGPPGPHGG--PGGPVGRPLGPPP--PRGDPFWDDPG--DPRGGPMRG 709  
QY 158 GPGPGGPMGR-MGCRGDRGCGFPFRGSRGNPSGNNVQHRAG-----DM 204  
DB 710 GPGGCGPYRRGCGNGNPPPPPPFRGARGRGSGPPNRGGPGCGGVMGGGHRPH 769  
QY 205 QCEPFGCGNCFAMRTECNQCKAPR-EGFLPPPPPPGDRGRG-----GPGMRGGRG 258  
DB 770 EGGCGGWNSS-----GHRPHEG-----PGGGMGSGHRPHEGPGSGGCGG 810  
QY 259 GLMDRGPGMFRGGRG-----GDRGFR--GGRGMDRGGRGGR--RGPP--GG 302  
DB 811 GHRPHEGPGGGISGSGHRPHEGPGCGMGAGGHRPHEGPGSGSGGHRPHEGPHGG 870  
QY 303 PPGPLMEQMGRRG-----GRGPGKMDKGEHR 330  
DB 871 PHGHRPHDVPHGRHGRHGRPPHHRHGRHGRH--GGGHR 909

RESULT 15  
US-09-949-016-6136  
; Sequence 6136, Application US/09949016  
; Patent No. 6812319  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: C1001307  
; CURRENT APPLICATION NUMBER: US/09/949, 016  
; PRIOR FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 6136  
; LENGTH: 1603  
; TYPE: PR  
; ORGANISM: Human  
US-09-949-016-6136

Query Match 13.8%; Score 265; DB 4; Length 1603;  
Best Local Similarity 29.6%; Pred. No. 2,2e-13;  
Matches 113; Conservative 14; Mismatches 125; Indels 130; Gaps 19;

QY 1 GGRGMSGAGER-----GCFNKPGPMDEGPDLDGPPVDPDESDNSAIIYVQGLNDV 54  
DB 1104 GERGYTSAGKSGRPGSGGLPGPPAGPRGERGPGQNSGKGD-----QGFO--- 1154  
QY 55 TLDDLAPFKOCGVVANKKRGQPMIHYLDKETGPKGDATVSYEDDPTAKAAVEMFDG 114  
DB 1155 -----GDP-----GFTGTPSGPFGKVGSPGPPGQAE----- 1183  
QY 115 KDFG-----SKLVSLARKKPPMNSMRGSLPPREGRMPPLRGPGGPGGPMGRMG 170  
DB 1184 KGSREGIRGPGSLPSPGPPGPGIQRAGLDGLDGKDGKPGLRGDP--GPAGPPGLMGP 1242  
QY 171 GRG-----GDRGPPPRGPRGSRGNPSGNNVQHRAGDMQCPNPGC--GNQNFAMRTE 221  
DB 1243 FKGTGHPGLPGPKDCCGPPGPGSTGRPAEG-----EPGAMGPG----- 1284  
QY 222 CNQCKAKRPPGFLPPPPP--PG-----GDRGGRGPGMRG----- 255  
DB 1285 -----RQPPGPHVPPGPPGPPAGISAVGLKDRGATGGRGLAGLPQOPGPPHGP 1339  
QY 256 ---GRGGLMDRGPGGM--FRG-----GRGDRG--GFRGGRGMDRGFGGGRGPP 300  
DB 1340 GEPGTDAAGKGGPPGKQFYGPFGPKDPPGAAGCKGAGKGRAGMP--GGPGKSGSMGP 1398

QY 301 GGPPLMEQMGRRGGRGPG 322  
DB 1399 VGPGR-----AGERGHPGAPG 1415

Search completed: February 18, 2005, 15:31:22  
Job time : 47 secs